Bacterial growth mechanisms and their role in cell size homeostasis and senescence

by

Maryam Kohram

B.S., University of Tehran, 2012

M.Sc., University of Akron, 2015

Submitted to the Graduate Faculty of the

Dietrich School of Arts and Sciences in partial fulfillment

of the requirements for the degree of

Doctor of Philosophy

University of Pittsburgh

2021

UNIVERSITY OF PITTSBURGH

DIETRICH SCHOOL OF ARTS AND SCIENCES

This dissertation was presented

by

Maryam Kohram

It was defended on

April 1, 2021

and approved by

Zoltan N. Oltvai, Professor, Department of Pathology and Laboratory Medicine, University of Rochester

Xiao-Lun Wu, Professor, Department of Physics and Astronomy

Ayres Freitas, Associate Professor, Department of Physics and Astronomy

Erik S. Wright, Assistant Professor, Department of Biomedical Informatics

Thesis Advisor/Dissertation Director: Dr. Hanna Salman, Associate Professor, Department of Physics and Astronomy

Bacterial growth mechanisms and their role in cell size homeostasis and senescence

Maryam Kohram, PhD

University of Pittsburgh, 2021

Growth is a fundamental feature of living organisms, which plays an important role in maintaining cellular characteristics such as cell size and contributes to cellular fitness. It is determined by the rate of biochemical reactions, their efficiency, and their collective organization in a cell, and it is strongly influenced by the environment and the nutrients available to the cell. This complexity can lead to significant fluctuations in the cell's growth rate. To avoid the accumulation of fluctuations over time and prevent processes from diverging, cells utilize control mechanisms to ensure the stability and accuracy of growth. The aim of this study is to better understand the dynamical processes and control mechanisms of cellular growth in the simple model organism *E. coli* bacterium, and how they contribute to cell size homeostasis and cellular senescence. We use an experimental setup consisting of a microfluidic device designed to trap single cells while continuously growing, to acquire long-term single cell measurements of cell-size and protein content. We apply new regression analyses methods to the measured growth dynamics, which are not motivated by any preconceived growth control model. Our results reveal dependencies among measured cellular variables that were not considered before, and which point to new growth control mechanisms. By further investigating these dependencies, we find that DNA concentration plays an important role in determining bacterial growth rate, which works to compensate for size differences acquired during cell division and thus allows cells to maintain size homeostasis. We then turn our attention to the effect of metabolism efficiency on growth rate and we study how various metabolic defects affect the ability of cells to grow in different

environments. We focus on how these defects affect cellular senescence and contribute to cellular death. Our results uncover aging effects and distinguish different phenotypes of cell death.

Table of contents

List of tables

List of figures

Preface

I would like to thank all of those who have helped me along this journey. First and foremost, I would like to express my sincere gratitude to my advisor, Dr. Hanna Salman. He has been my role model in my path to becoming a physicist, and in setting my path for the future. He created an inspiring environment for creativity, encouraged me to test out my theories, and equipped me with an array of technical skills. His limitless support and valuable insights in both academic and personal matters have been the light to my darkness.

Besides my advisor, I would like to thank those who have helped along the way in contributing to my research. Thank you to Dr. Zoltan Oltvai for the long hours of meetings and discussions, his perspective was and still is a huge contribution to my research on bacterial aging and senescence. Dr. BingKan Xue and Dr. Stanislas Leibler were our collaborators on the regression research; thank you for your ideas and your huge contribution to the data analysis that were obtained from our lab. Thank you to Dr. Naama Brenner for her impactful discussions with our group and her help in shaping the theories in my research. I would also like to thank my dissertation committee, Dr. Wu, Dr. Wright, and Dr. Freitas for their guidance, encouragement, and support throughout my graduate studies. I want to thank (soon to be Dr.) Harsh Vashistha; we have worked together in the lab for the past four years and he has helped me learn and grow to become a scientist. I will always remember our hour-long debates on science and philosophy. Dr. Zhicheng Long, thank you for teaching me lab techniques used in the clean room and for guiding me in my way to become an experimental scientist.

Of equal importance is the love and support I have received from my family and friends. Dad, your contributions to science were the inspiration behind my choice to become a scientist.

You always gave me advice and emotional support and reminded me of how proud you are of me. So, thank you, you are the reason I have come this far in my life, and will continue my path as a scientist. Mom, thank you for always being there for me, helping me through hard times, and encouraging me to succeed. I grew up watching you getting your PhD while holding a job and caring for a family; you have inspired me to work hard to achieve my goals in life. There are no words that can explain my gratitude to my brother, my sisters, and my boyfriend. Mojtaba, thank you for always being there for me and being my rock, especially when I had just arrived in the US. I wouldn't have survived those first years without you. Fatemeh, thank you for your endless support in my work and life. You are the big sister that everyone wishes they could have. I still know I can run to you with all my problems, and you will always be there. Zahra, thank you for being the one person in my life that I can talk to about anything and everything, and know that you will understand me without any explanation. I can always count on your support. Behnam, you are the love of my life. Thank you for your boundless love, encouragement, and support. Your presence has made this journey more enjoyable. Yasemin and Wendy, thank you for your encouragement and endless support through all the ups and downs on my way to a PhD. I can't imagine the past five years without your love and support.

This dissertation is dedicated to the memory of my father, Hamid Kohram. You have inspired me to pursue science.

1.0 Introduction

Growth is a fundamental requirement for self-replication to maintain all life forms, and it is the integrative result of many cellular processes. Steady and reliable growth is achieved by cellular homeostasis, which is the regulation of internal processes to enable living cells to maintain their physical and functional characteristics even when facing significant changes in their environment. Examples of homeostasis in human beings include blood sugar levels, body temperature, or breathing patterns, which are fundamental for health; These properties are maintained within a constant range despite changes in diet or changes in the weather[1,2,3]. How physical and chemical homeostasis are preserved in various organisms and how these homeostasis mechanisms control individual growth has been one of the fundamental questions in biology.

Nevertheless, with all the homeostatic mechanisms involved, living cells will deteriorate over time as they grow and divide to produce offspring, a process known as senescence, which eventually leads to their death. How cellular growth progress and contributes to cellular senescence in all living organisms has been another fundamental question in biology for decades. Despite extensive studies in recent times aimed at understanding the mechanisms controlling growth[4-5] and aging-related senescence[6-7] in bacteria, we are still unable to determine all molecular mechanisms that influence these processes and their variation among individual cells within a population experiencing identical environmental conditions. My research is dedicated to the understanding of how growth is controlled in bacteria, and how it contributes to cell-size homeostasis and senescence.

Over the years, different mechanisms for bacterial size homeostasis have been proposed. The three major models that have been extensively studied are the “sizer”, “adder”, and “timer”

(Figure 1). The sizer model[8,9] proposes that each cell monitors its size and divides when it reaches a specific threshold size. In this mechanism, as the birth size of a cell becomes larger, the amount of added size during the following cell cycle decreases, concluding in the negative slope presented in Figure 1A. As a result of cell division occurring at constant size, and that *E. coli* cells are divided symmetrically, the birth size of a cell, in this model, is expected to converge to the steady state size within one generation (Figure 1B).

In contrast to sizer, the adder model[10–12] proposes that a single cell adds the same amount of size in every cell cycle regardless of its birth size (Figure 1A). This strategy would lead to the birth size converging to its steady state value within less than two generations[12,13] (Figure 1B).

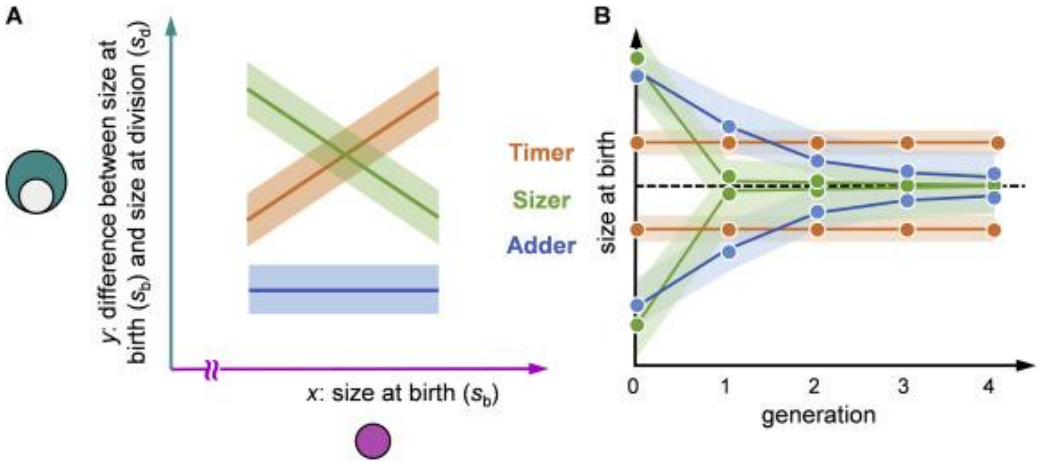


Figure 1 Three behaviors for size control. A) When added size during a cycle does not depend on birth size, the mechanism is named adder, when they are correlated with a slope of -1 it is sizer, and when they are correlated with a slope of +1 it is timer. B) The effect of the three size homeostasis behaviors on cells. The dashed line presents the steady-state. Figure adapted from Facchetti et al., 2017[8].

The timer model on the other hand, proposes that each cell can measure time, and it attempts to grow for a specific amount of time before division. The generation time in this model does not depend on birth size. Since the time that a cell grows during a cycle is constant, as the birth size increases, the size that a cell adds during that cycle also increases, leading to the positive slope seen in Figure 1A. With this mechanism, the birth size diverges from the steady state size

(Figure 1B), and therefore it cannot account for maintaining size homeostasis independently. Nevertheless, some studies have suggested the existence of this mechanism for controlling homeostasis of other factors in the cell such as protein folding[14].

Note that the above-described models of size control are phenomenological and are not based on any known molecular mechanism. More recently however, a mechanistic model that provides molecular basis for the adder phenomenon has been proposed[15]. The cell cycle in this model is divided into three major intervals designated as B, C, and D[16]. B starts right after cell division until chromosome replication initiation, C is from the start until the end of replication, and D is from the end of chromosome replication until cell division. This coordination of cellular events reveals a correlation between cell size at replication initiation and cell size at birth, and a link between growth initiation and DNA replication that accounts for the adder phenomenon observed in many bacterial species.

Traditional experimental methods, such as agar plates or liquid cultures for measuring cellular growth, are unable to fully capture and characterize the growth dynamics of individual cells. On the other hand, having a comprehensive understanding of the heterogeneous nature of bacterial cells, which allows cells with the same genetic code to have variable properties even when in the same environment, is necessary for developing detailed description of the regulation mechanisms underlying fundamental biological processes such as growth and division. This requirement has led to developing a new experimental method, in which the growth of single cells could be measured accurately. The new method, which makes use of a microfluidic devices that allow for single cell trapping and monitoring for extended periods of time (see Chapter 2.0), has been used in recent years to investigate size homeostasis in bacteria, providing high-throughput, high-quality, single-cell level measurements of cell-size dynamics. As a result, researchers have

been able to gain a better understanding of how a cell changes in size and composition over time, and how cellular growth proceeds, and ultimately develop mathematical frameworks to model these changes.

However, previous studies have been limited in several aspects. Firstly, the dynamic range of cellular growth does not allow for clear distinction of the mode of growth. The best fit for the measured dynamics has been simple exponential. Thus, cellular growth has been assumed to follow an exponential growth curve, whose exponential growth rate fluctuates between cell cycles. Recently, this purely exponential growth has been challenged. Nordholt et al.[17] argue that *Bacillus subtilis* do not follow a constant exponential growth program and that they observe a biphasic dynamic[18]. Secondly, examination of the measurements in previous studies have been guided by preconceived models, which predict certain correlations and/or dependencies that are then tested in the data. This can often lead to missing correlation that are not accounted for in the model being tested.

In this study, we use an agnostic approach in examining the dynamical data of cell size (see Chapter 2.0). Our goal is to be able to find how the growth characteristics of a cell is inherited, or how it affects the properties of its daughters, and following generations. We measure different properties of growth such as cycle time, birth length, and growth rate, and examine the relationship between them in consecutive generations. We find that several correlations exist simultaneously among many cellular properties, which suggest that the dynamics of cell growth and division is determined by multiple mechanisms whose integrated actions give rise to the homeostasis[19,20]. Our results recover correlations between cellular properties that have been previously reported for the adder model[12], but more importantly, they reveal additional dependencies that were never discussed before. One newly revealed dependency between the growth rate and the size fraction

points to a feedback mechanism, which could potentially explain how cells control their growth rate in order to make up for size differences acquired during cell divisions – cells that receive a smaller fraction of the mother cell tend to grow faster than their sisters. Similar mechanisms have been reported in different cell types and recent studies have provided evidence that suggest individual animal cells are measuring and adjusting their growth rate.[21]. The growth rate of animal cells was shown to increase as the cell size decreases[22]. To further study the relationship between sister cells, we have measured the growth dynamics of sister cells immediately after division and throughout the first cell-cycle following their separation. Besides validating the new growth control mechanism, we find additional correlations between cellular properties of sister cells and genealogically more distant cells that were unaccounted for by previous models of cell-size control based only on single cell data.

Following these findings, I initiated a study that aims to explore the source of this growth rate asymmetry between sister cells in order to better understand the control mechanism of cellular growth in general and how it contributes to size homeostasis. To gain a better understanding of this, we need to direct our attention towards the molecular content in a bacterial cell and understand what processes govern cellular growth and division. In reality, thousands of molecules are involved in cellular growth and division and many biochemical reactions are required, but here we make an effort to simplify the model of the cell as much as possible by categorizing the main chemical components of a dry *E. coli* bacterium cell into five groups: DNA (3%), RNA (20%), proteins (50-55%), lipids (7-9%), and carbohydrates, soluble metabolites, and salts (16-18%)[23]. Note that this is an over-simplification of the chemical composition of a cell and a more thorough table can be found in *Escherichia coli and Salmonella: cellular and molecular biology* (Dennis and Bremer, 1996)[24]. However, this classification is sufficient for the purposes of this research.

The size and composition of a cell are sensitive functions of growth rate and previous studies have been conducted to find the relationship between the rate of synthesis of these molecules and their amount in a cell with respect to growth rate. Measurements made in cultures of *E. coli* B/r growing exponentially at 37℃ in different growth media revealed that the values of RNA-to-protein and DNA-to-protein ratios are increasing and decreasing, respectively, with an increase in growth rate. This study by Dennis and Bremer also reports the number of replication origins per cell at initiation to be 1, 2, 3, or 4, with higher growth rates having a higher number of origins[24]. In another study, Cox found that RNA-to-protein ratio is a key property of *E. coli* bacterial cells, and that their specific growth rate is directly proportional to $m_{RNA(ave)}/m_{p(ave)}$, where $m_{RNA(ave)}$ and $m_{p(ave)}$ are the average RNA and protein masses of one cell, respectively[25]. Note that all parameters reported in the two mentioned studies are averaged over many cells in a population and do not provide any information about the phenotypic heterogeneity among single cells in the same population. More recent studies[26] have investigated the differences in the amounts of chemical composition of cells during single cell division and growth by considering the role of molecular stochasticity of metabolic reaction events and the expression level of metabolic proteins. Kiviet et al.[27] showed that the instantaneous growth rate of single cells of *E. coli* can fluctuate with the expression fluctuations of catabolically active enzymes, and that noise transmission depends on the limitation of the enzyme to growth. Zhang et al.[5] changed the concentration of free RNAP in single *E. coli* cells in an effort to decrease the amount of RNAP available for transcription of genes, and concluded that this is the main factor limiting growth rate of single cells. Several similar studies have been conducted in an attempt to discover the main effectors of growth rate in *E. coli* bacteria. A high-throughput analysis of metabolic enzyme genes disruptants revealed that *E. coli* cells regulate their metabolic enzymes levels in order to keep their metabolic state robust with

changing growth rates[28]. Another study on the ribosomal components and the rRNA processing in *E. coli* suggests that growth rate control is directly related to the synthesis of rRNA[29].

The effect of cellular composition on the cell's growth rate, in addition to the fact that the smaller cell grows at a faster rate than its larger sister, suggest that at least one of the resources that influences the cell's growth rate is divided unequally between the sisters such that it causes the smaller cell to grow at a faster rate. This could be a growth "activator" present at a higher concentration in the smaller sister cell, or a growth "inhibitor" present at a lower concentration. Since no growth inhibitor has been identified in bacteria, we consider the first possibility. We propose that the nucleoid, which is divided between the two sisters in equal copy numbers and thus would have a higher concentration in the smaller cell, might play an important role in determining the cell's growth rate. This possibility is supported by the fact that all proteins and RNAs in the cell are produced from the DNA. Therefore, the rate of reactions that produce these materials, which are required for the cell to grow, depend on the DNA concentration. Several new investigations that have been reported recently support this possibility[30-31]. We further test this hypothesis in Chapter 4.0 and show that indeed, as the DNA concentration in the cell increases, so does the growth rate.

Finally, we turn our attention to the effect of growth on aging in bacteria. In mammalian cells, aging is defined by the inability of the cell to carry out the physiological functions necessary for survival[32]. However, in single cell organisms, such as bacteria, there has been various definitions. Over the last 60 years, some unicellular species were considered to be immortal due to their symmetrical division. Nevertheless, biologically immortal organisms can die due to factors other than senescence, such as injury or disease[33]. Recent studies indicate that in various tissues, cells that enter into a non-dividing state and senescence-associated secretory phenotype are major

contributors to the onset of aging-related diseases[34]. Until now, aging-related processes have been mainly studied in eukaryotic organisms. Researchers have used various eukaryotic model organisms to identify cellular pathways that contribute to aging-related processes, from single-celled species to simple multicellular animals and higher vertebrates[35,36]. *Saccharomyces cerevisiae* is a species of yeast (budding yeast), which despite being a unicellular organism, has been a long-standing platform for studying senescence due to its asymmetrical division that uniquely distinguishes between mother and daughter. This type of division concludes with the mother cell sustaining all the damage, while the daughter cell is rejuvenated. In *S. cerevisiae*, aging has been measured in two distinct periods. The first is "*replicative life span*" (RLS), counting the number of cell divisions a budding yeast cell can undergo before entering a permanently non-dividing state, termed "replicative senescence", or in other words, the lifetime before the budding ceases. The second is *chronological life span* (CLS), i.e., the time span that permanently non-dividing, replicative senesced cells remain viable. Unbiased genetic screenings in *S. cerevisiae* have identified critical determinants of RLS[37,38]. However, similar studies in prokaryotes have been uncommon though their results could contribute to the development of a unified theory of cellular senescence. Recently it has been shown that even symmetrically dividing unicellular organisms, such as *E. coli*, which were considered to be immortal, are revealing signs of aging[39,40].

While a specific definition of aging in symmetrically dividing bacteria is not universally agreed upon, most recent studies take either one, or both of the following two factors into consideration when addressing replicative senescence: growth rate[39–42], and survival function[6,40,43]. The first factor considers the cell as aging if its growth rate reduces over time, and as immortal in the case where its growth rate is stable over time. The second factor uses the Kaplan-Meier survival function estimate[44]:

$$S(t) = \prod_{t_i \leq t} [1 - \frac{d_i}{n_i}]$$

where $S(t)$ is the survival probability function, t is time, d_i is the number of deaths at point i, and n_i is the number of individuals at risk of death prior to point i. In different studies, $S(t)$ has been fitted to a number of distributions. Currently, only two forms of this function have been reported in studies of eukaryotic senescence. The first is a simple exponential decay[6]:

$$S(t) = e^{-at}$$

where $a > 0$, and is considered to be the rate of death. The probability of cell death during a certain time is called the hazard function (or hazard rate, or failure rate, $\lambda(t)$) and can be calculated using $\lambda(t) = -\frac{dS/dt}{S}$ [45]. In the exponential limit, $\lambda(t) = a$, and thus the hazard function does not depend on time, indicating age-independent or stochastic death in the population[46]. A population displaying this behavior is considered to exhibit no aging. The second behavior can be fitted to a Gompertz distribution[47]:

$$S(t) = e^{-\frac{a}{b}(e^{bt}-1)}$$

Where the death rate $b < 1$, since a rate of 1 or higher would lead to the extinction of the population. The hazard function is then $\lambda(t) = ae^{bt}$. This hazard function increases with time, indicating increased death probability with time, which signifies aging. A population exhibiting this behavior is considered to be undergoing aging[48].

The theory of immortality in unicellular species was first challenged in 2005, when Stewart et al.[39] reported a change in the growth rate of *E. coli* cells with age. This study differentiated between the two poles of the cell, since one of the poles is always created anew during the cell's last division, while the other pole is not renewed and thus it determines the age of the cell. The age

of the cell then is defined as the number of divisions the old pole of the cell has survived. Using this designation, the two supposedly similar cells are functionally asymmetric due to their different age[39]. The results of this study showed that the cells with the old pole exhibit a decrease in growth rate relative to their younger sisters as they age. In 2010, Wang et al.[49] obtained results in contradiction to Stewart et al., indicating the growth rate of a cell is robust. This contradiction was explained by the experimental limits of Stewart et al. methods, which allowed them to follow a cell up to only five generations. Wang et al. also plotted the fraction of surviving cells as a function of time, showing the function is not an exponential decay, indicating that there are age-dependent factors in the population death. In recent years, other studies have demonstrated a change in the growth rate of *E. coli* cells only in presence of stress, such as antibiotics, mutations, and heat[7,50–52]. These studies agree that *E. coli* cells exhibit no age-related change in the growth rate in the absence of stress.

The question then is how is the age of the cell manifested in its properties such as growth rate or accumulation of molecules, and how does it lead eventually to death in the absence of stress? One possibility is that aging reduces the efficiency of how cellular processes are performed. As an example, it could lead to reduced efficiency of the metabolic activity of the cell, yet without influencing the cellular growth rate. To address this question, I took a different approach, by asking: can the metabolic efficiency influence the cell's senescence without influencing the cell's growth rate?

Hottes et al.[53] were able to demonstrate multiple loss-of-function mutations that were beneficial to cells in changed environments. Takeuchi et al.[54] took advantage of their newly developed high-throughput method to measure colony growth of a set of 3985 single-gene knockouts that were created by Baba et al. (the Keio collection)[55] with the objective of easier

analyses of unknown gene functions. Their results indicated that many mutants grew similar to wild type cells on an agar plate, and some mutants exhibited defective growth. Conversely, these defective grown colonies might show a metabolic benefit in a different environment or affect the process of aging in the cells.

By taking advantage of this mutant collection, we identified a specific mutation that does not change the growth rate of the cell under certain environmental conditions. Nevertheless, we find that such mutation does reduce the metabolic efficiency of the cell and increases its senescence (see Chapter 5.0). Our results confirm that deletion of the α-subunit of the ATP synthase increases cellular respiration[56]. A thorough examination of this mutation at the single-cell level in the microfluidic traps reveals different phenotypes of cell death, and that the fraction of phenotypes within the population is altered with the end result of accelerating cellular death upon deletion of the α-subunit of ATP synthase. Similarly, wild type cells display different phenotype fractions in a nutrient-limited environment.

2.0 Design and fabrication of microfluidic machines

Researchers have been studying populations of bacteria for several decades for different purposes such as understanding certain diseases, human health, and antibiotic resistance, which have resulted in significant outcomes such as the development of useful drugs[57,58,59]. However, bulk culture measurements cannot provide a detailed picture of cell dynamics, since phenotypic differences among individual cells within the population are not considered. Due to the exponential replication of bacteria in desirable growth conditions, studying one single cell is not a simple task and when on agar plates, only six to ten generations can be traced before the number of cells become too high and impossible to track[39]. Recently, studying single cells for hundreds of generations has become conceivable due to microfluidic machines.

A microfluidic machine is a device that can control micro-scaled experiments due to its large surface to volume ratio. These devices are currently being used in many areas with promising results such as biological and biopolymer experiments, chemical analyses, microelectromechanical systems and optical telecommunication[60]. To produce such a machine, a permanent mold is first fabricated using photolithography. That mold is then used to prototype microfluidic devices that are made from elastomers, mostly polydimethylsiloxane (PDMS), due to its unique physical properties that make it useful in many areas. PDMS is a viscoelastic elastomer, meaning that at certain temperatures it acts as a viscous liquid, but at other temperatures it acts as an elastic solid. It has the property of being highly permeable to gases, making oxygen constantly available inside the machine[61]. PDMS is also optically transparent, non-toxic, biocompatible and cheap in cost. It is hydrophobic, but when treated with plasma oxidation the surface becomes

hydrophilic temporarily, giving it the ability to bond with other layers to produce a useful machine[62].

Using such devices, has made trapping single cells for extended times, while washing away daughter cells to prevent accumulation of bacteria, feasible and simple. These devices also allow creating a homogeneous environment by constantly feeding the trapped bacteria. However micro-niches can still exist within the device. One example trapping method that was recently developed is to use a strain where surface adherence features are integrated at a single chromosomal locus under the control of an inducible promoter[63]. After sufficient number of cells are stuck to the surface of the device, the excess cells can be washed away, and the inducer is removed so that no other cells adhere during the remaining of the experiment. In this research (chapters 3.0, 4.0, and 5.0), another microfluidic device named the "mother machine"[40,64] has been used to study growth and replication of single cells. This chapter is dedicated to detailing the design and fabrication of such machines, while the methods of utilizing them for our specific purposes are explained in each chapter.

For my research purposes, I used two different photolithography methods to fabricate the same initial mold, namely mask alignment (Section 2.1.1, Figure 3) and 3D lithography (Section 2.1.2, Figure 12). Once the mold was generated, the PDMS replicas were always prepared with the same approach (Section 2.2). A comprehensive explanation of the steps for each photolithography is presented in the following sections, followed by the method used to prepare the PDMS replicas.

2.1 Fabrication of permanent mold

2.1.1 Mask aligner

In this method, a contact mask aligner (Quintel Q4000 MA) represented in Figure 2 was used to create the mother machine. This system allows accurate alignment of semiconductor wafers with a mask and exposes ultraviolet light to create the pattern of the mask on the wafer. The mother machine consists of channels with two different heights, which mandates a two-layers fabrication procedure. The process of how each layer was fabricated is presented in Figure 3.

Figure 2 Quintel Q4000 MA Mask Aligner located at Nanoscale Fabrication & Characterization Facility (NFCF) at the University of Pittsburgh. Picture adapted from NFCF webpage.

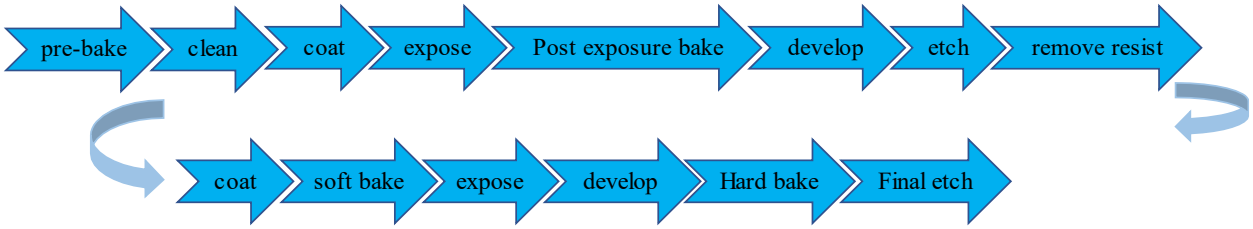


Figure 3 First line is the process of first layer fabrication, and second line is the process of the second layer.

First, the channels with a lower height were fabricated on the substrate, and then the second layer with channels that were much higher was aligned and fabricated on top of the first layer. Hence, two masks had to be created in the beginning. These masks were designed with the desired pattern in AutoCAD® and were then ordered from Photo Sciences, Inc. (Torrance CA). Figure 4 shows a sample design of the mother machine in AutoCAD® with separated layers. Layers had to be designed according whether a positive or negative photoresist was used (see next paragraph for more detail). The substrate that was used for mask alignment was a 3-inch silicon wafer. Figure 5 presents a complete view of the two layers of the machine in a larger scale.

The first layer of our design consists of small channels to trap single cells. These channels were designed to be 1 μm wide and 1 μm high so that the cells would have a hard time escaping from them. The length of the channel was designed to be two different sizes, 15 μm and 30 μm. For the effect of channel width and length on cell growth see reference [65]. To achieve these features, I used a positive resist, S1805. Figure 6 explains how photoresists work, and the difference between a positive and a negative resist. A positive resist is a chemical that binds very well to the substrate, and when exposed to light with specific wavelength, the binding becomes weak. This means that during the development step, all parts that were exposed to light will be washed away. In our case the small channels were not exposed to light, while everything else was. Accordingly, the AutoCAD® design was such that passage of light was blocked where the small channels were intended.

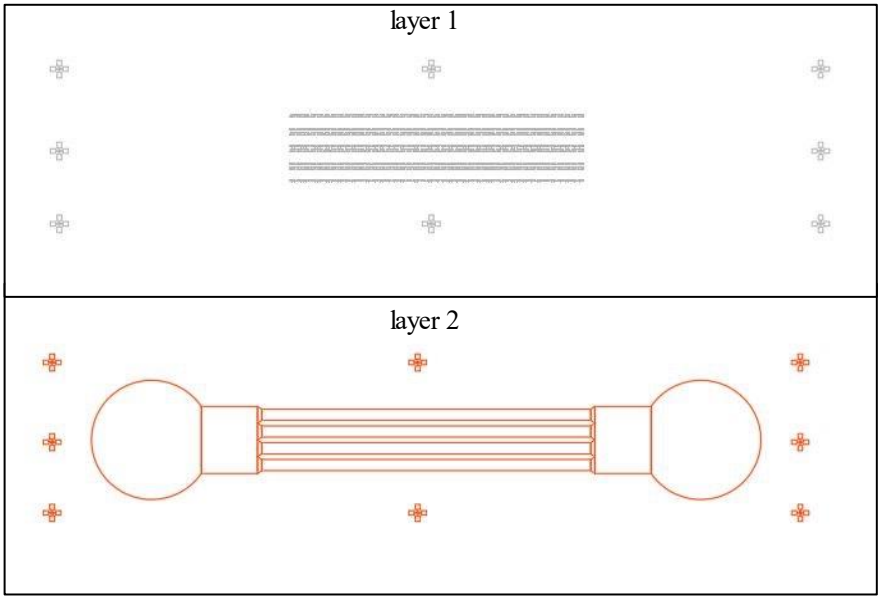


Figure 4 Two masks designed in AutoCAD® for the mother machine. Layer 1 are the small channels (traps) and layer 2 are the large channels.

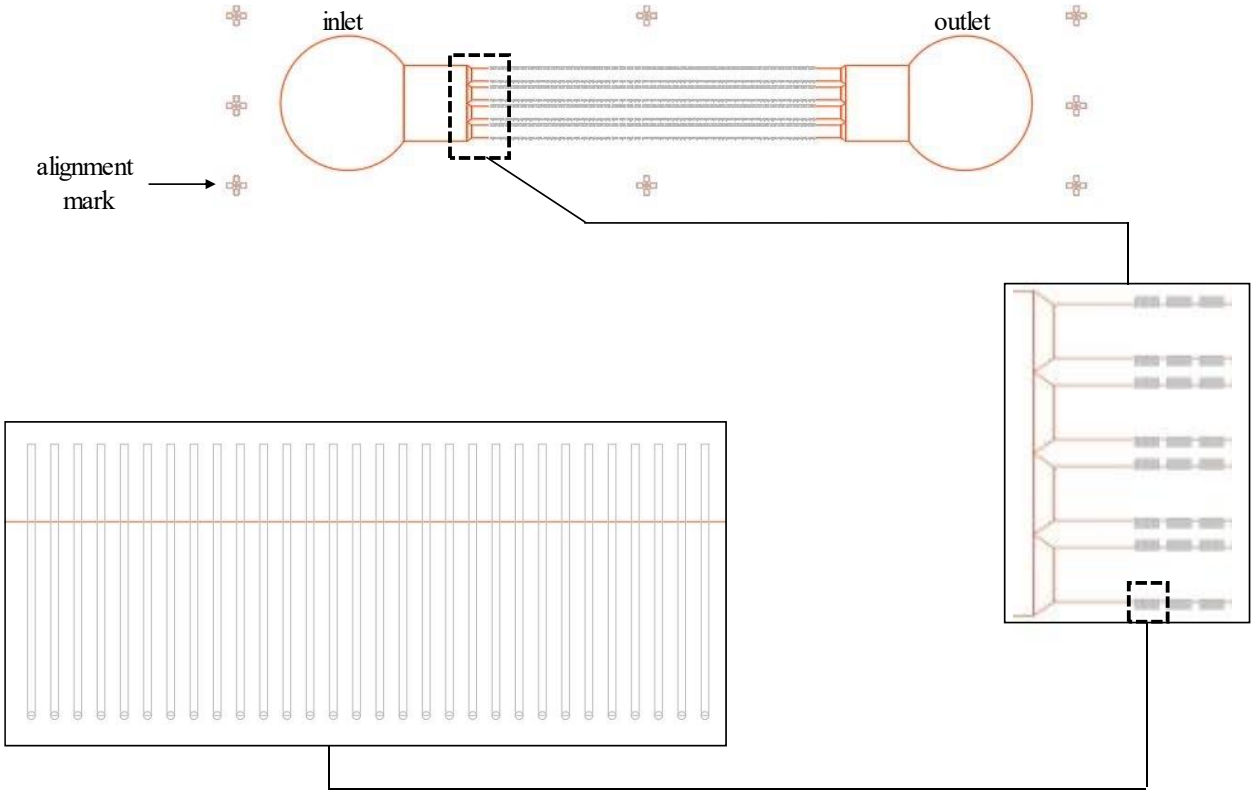


Figure 5 A complete look at the desing of the mother machine in AutoCAD®. Small channels (layer 1) are black and large channels (layer 2) are orange.

The second layer of our device consists of large channels with 30 μm height and 1 mm length. For these dimensions, I used the negative resist, SU8-2015. A negative photoresist is a chemical that binds to the substrate temporarily, and when exposed to light with specific wavelength, the binding becomes very strong (Figure 6). Thus, during the development process, the substrate that was not exposed to light will be washed away. Accordingly, our AutoCAD mask design allowed passage of light where the large channels were intended.

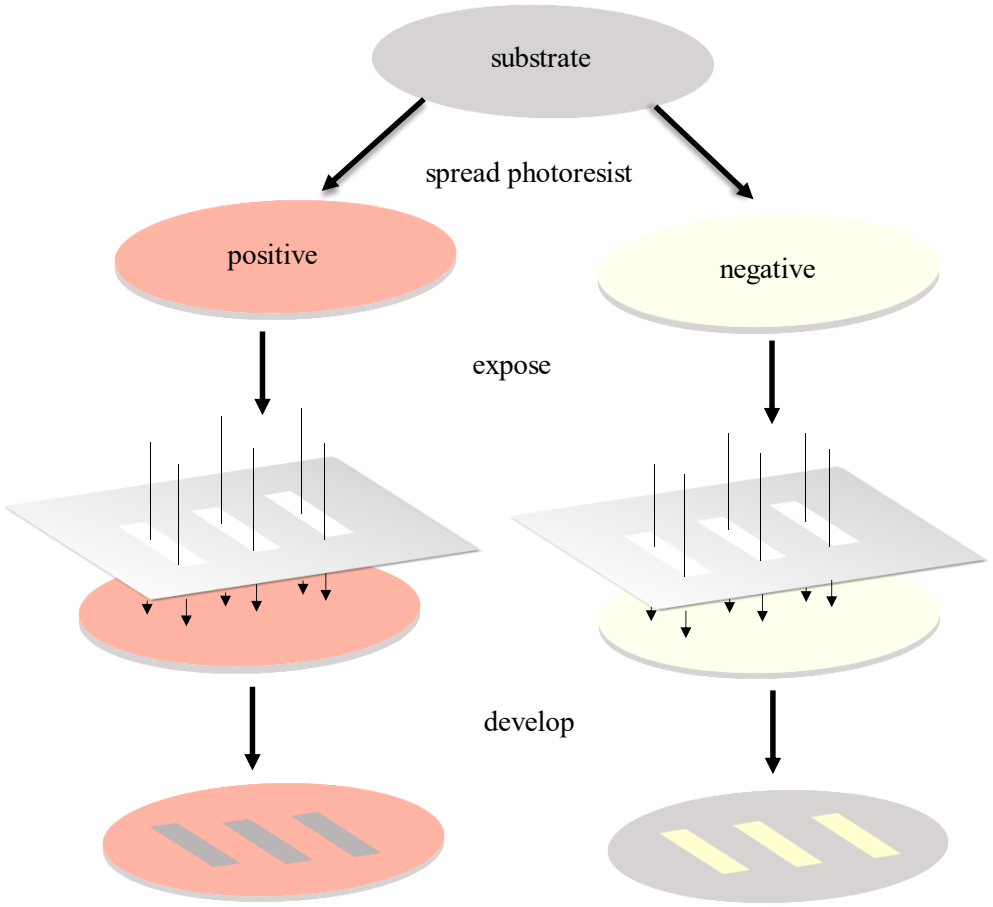


Figure 6 Visual explanation of positive and negative photoresist. When a positive photoresist is developed, the parts that were exposed to light are washed away, but when a negative photoresist is developed, parts that were not exposed to light are washed away.

2.1.1.1 First layer preparation

1- Pre-baked the substrate (3-inch silicon wafer) at 200℃ for 5 minutes by placing it on a hotplate (Barnstead Super-Nuova).

2- Cleaned the surface of the substrate using a Reactive Ion Etcher (Trion Phantom III LT RIE) with O_2 for 60 seconds.

3- Spread the first layer - S1805 optical - using a spin processor (Laurell WS-400B) with 2000 rpm for 50 seconds. This created a layer that is almost 1 μm high (Figure 7).

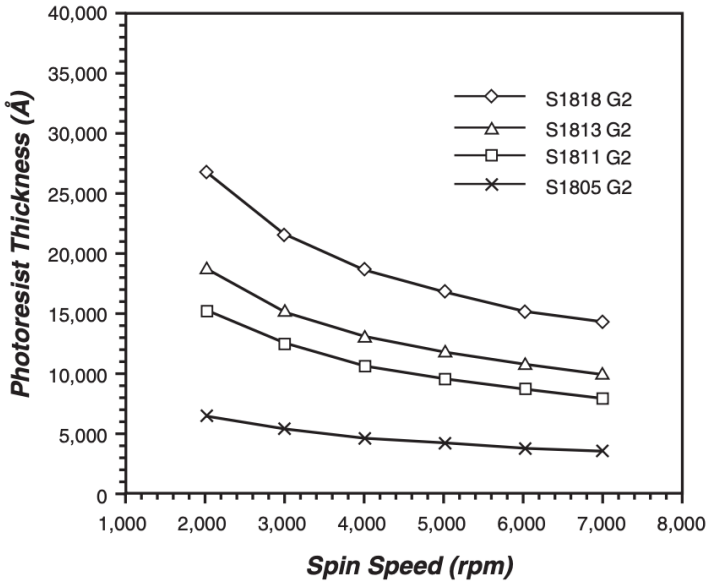


Figure 7 Spin speed curves of MICROPOSIT S1800 photoresist undyed series adapted from Rohm and Haas, 2006[66].

4- Soft baked the substrate at 110℃ for 1 minute by placing it on a hotplate.

5- Used Quintel Q4000 MA on vacuum mode with a wavelength of 365 nm and measured the lamp intensity.

6- Calculated the exposure time from Figure 8 (~20 seconds).

7- Placed the mask onto the mask holder of the mask aligner and selected vacuum contact mode. Then placed the wafer on the stage and aligned it with the alignment bar. Pressed the expose button. Unloaded mask and wafer from the instrument.

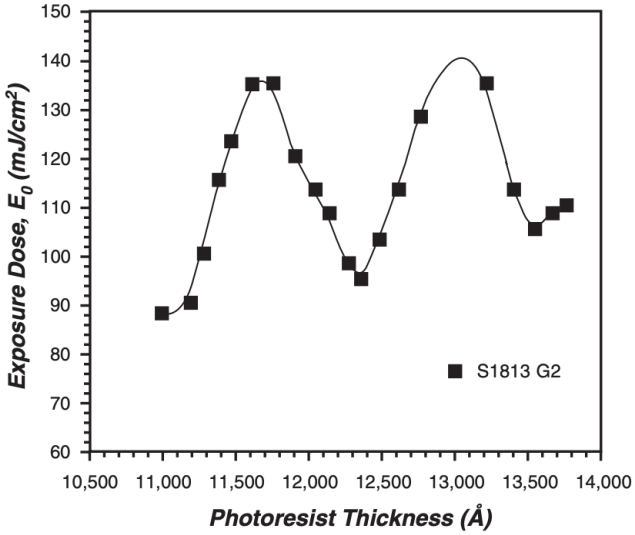


Figure 8 Photoresist interference curve adapted from Rohm and Haas, 2006[66].

8- Post-exposure baked the substrate at 115℃ for 1 minute by placing it on a hotplate.

9- Used 351 developer, mixed it with water at 15% final concentration, and used it to develop the substrate in it for 30 seconds.

10- Checked the substrate under the microscope to make sure the development is complete. If not, developed it for an additional 10 seconds or until completely developed.

11- Etched the substrate using RIE with the following recipe: $SF_6 = 25, O_2 = 4, CHF_3 = 10$, for 100 seconds. This step is a trial to to calculate the etch rate needed, since this rate is always variable.

12- Used a surface profiler (KLA-Tencor AlfaStep IQ Surface Profilometer) to measure the depth of the etched parts. Calculated the etch rate and put the substrate back into RIE for the required time.

13- Checked the height with the surface profiler at the end again. For *E. coli* cells the best height is exactly 1 μm.

14- To remove the photoresist, I made 3:1 Piranha mixture and warmed it to 150℃, then washed the substrate in it for 10 minutes.

15- Washed the substrate extensively with water afterwards to make sure residuals of the Piranha mixture has been removed.

2.1.1.2 Second layer preparation

1- Spread the photoresist SU8-2015 using the spin coater in two stages: 20 seconds at 800 rpm and 60 seconds at 500 rpm. This resulted in a layer $\sim 20 \mu m$ high (Figure 9).

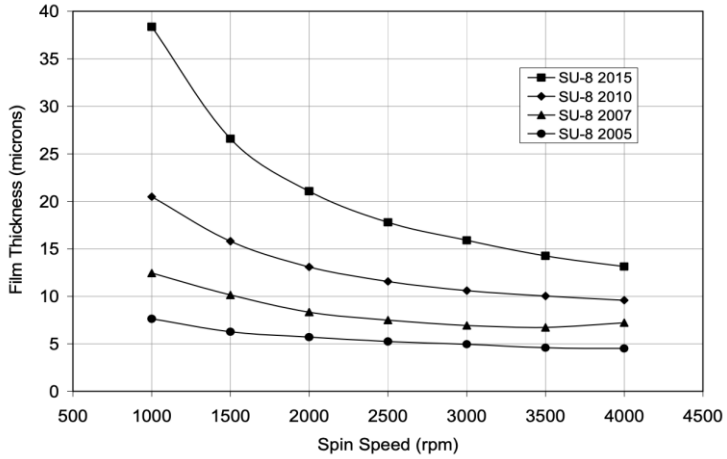


Figure 9 Spin speed vs. thickness for SU8 2000 adapted from MichroChem, 2015[67]**.**

2- Left the substrate in dark for one day to allow the photoresist-substrate binding to stabilize.

3- On the following day, I soft baked the substrate with the photoresist coating at 65℃ for 5 minutes, and then at 95℃ for 10 minutes by placing it on a hotplate.

4- The second layer mask was then aligned with the first layer using the mask aligner and predetermined marks on both layers.

5- The second layer photoresist was exposed to light (wavelength 365 nm) for 60 seconds. This exposure time was determined using Table 1.

6- The second layer was then developed using SU8 developer for 6 minutes.

7- The resulting structure was then inspected under the microscope to make sure that the development step removed all excess photoresist. In cases where excess photoresist was observed, a supplementary development step was applied for an additional 1-2 minutes or as needed.

8- Hard baked the final structure at 200℃ for 20 minutes.

9- Finally, I etched the complete structure using RIE with SF_6 for 30 seconds to acquire a smooth surface.

Following this procedure, the resulting mold was then ready for salinization described in (section 2.1.3).

Table 1 Exposure dose for SU8 2000 series adapted from MichroChem, 2015[67].

Thickness (μm)	Exposure Energy (mJ/cm^2)
0.5 – 2	60 – 80
3 – 5	90 – 105
6 – 15	110 – 140
16 – 25	140 – 150
26 – 40	150 – 160

2.1.2 3D lithography printing

This process also requires two separate steps; however, each step has fewer procedures compared to fabrication with mask alignment. The overall process of fabrication of each layer with

this method is displayed in Figure 12. The first layer was printed using a 3d lithography system (Nanoscribe Photonic Professional (GT) - Dip-in Liquid Lithography model (DiLL) high resolution) represented in Figure 10, and the second layer was fabricated using a maskless aligner (Heidelberg MLA100 Direct Write Lithographer), represented in Figure 11. The substrate used in this process was a 30x30x0.7 mm square fused silica. Both layers were designed in AutoCAD® similar to the method explained in the previous section. The design of the second layer was exactly the same as the one explained in section 2.1.1. We designed the first layer also the same as before and then made it 3d and exported it from AutoCAD® as an .stl file. A sample view of a couple of small channels is presented in Figure 13. This file was converted to .gwl at the end for the 3d printer to be able to read and print it.

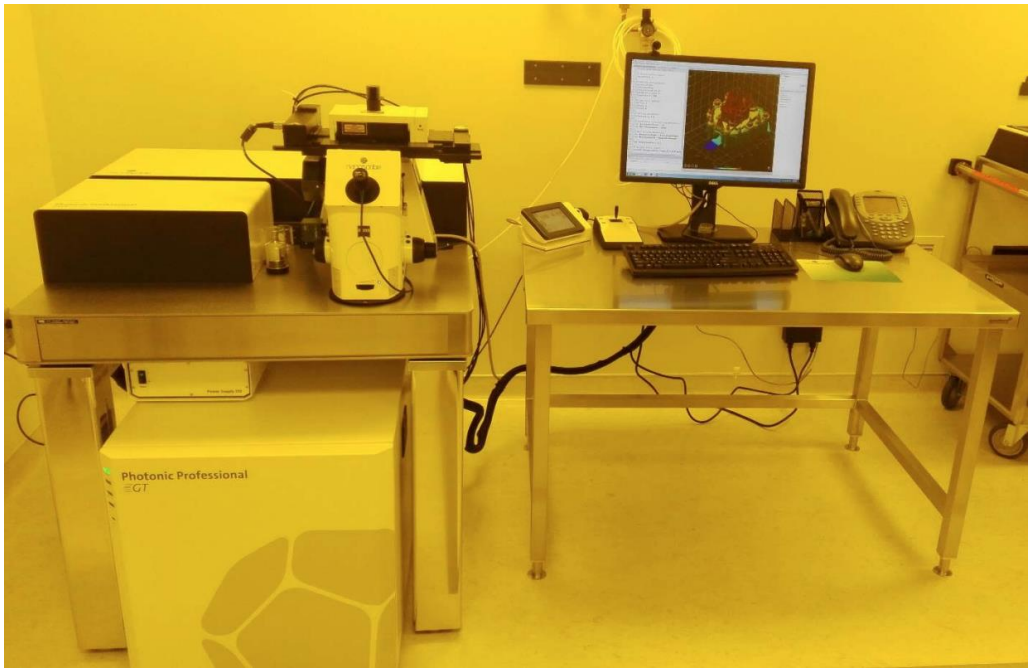


Figure 10 Nanoscribe Photonic Professional (GT) – Nano 3d Printer located at NFCF at the University of Pittsburgh. Picture adapted from NFCF webpage.

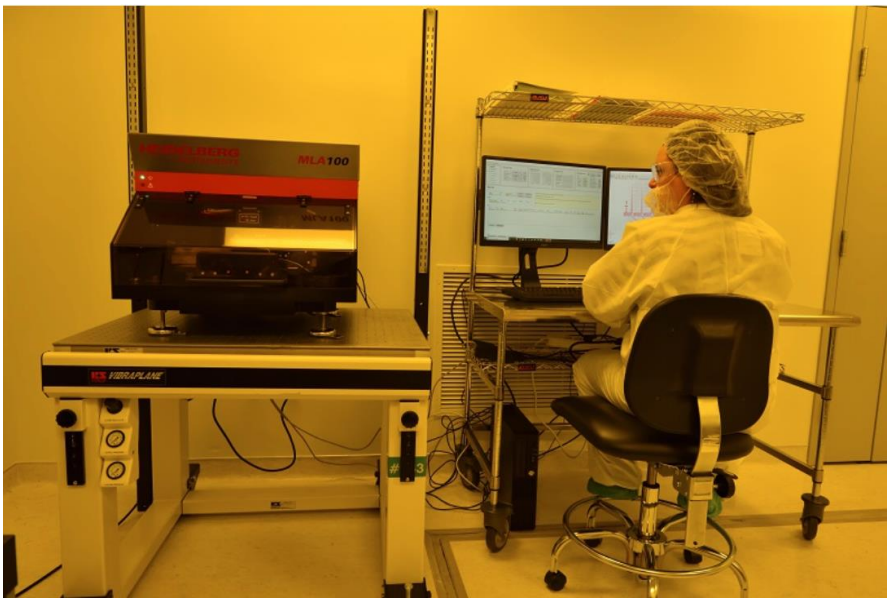


Figure 11 Heidelberg MLA100 Direct Write Lithographer located in NFCF at the University of Pittsburgh. Picture adapted from NFCF webpage.

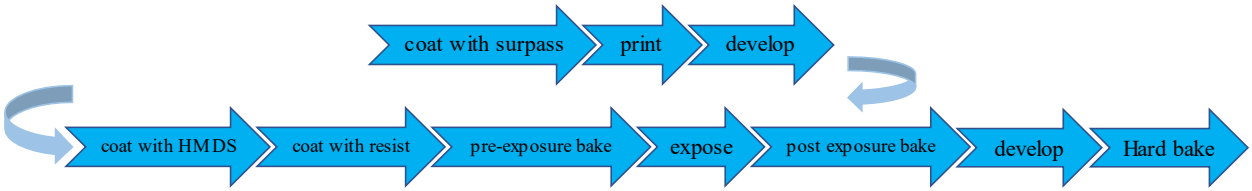


Figure 12 Process of fabrication of two layers of mother machine using 3d printing. First line represents fabrication of first layer (small channels) and second line represents the second layer (large channels).

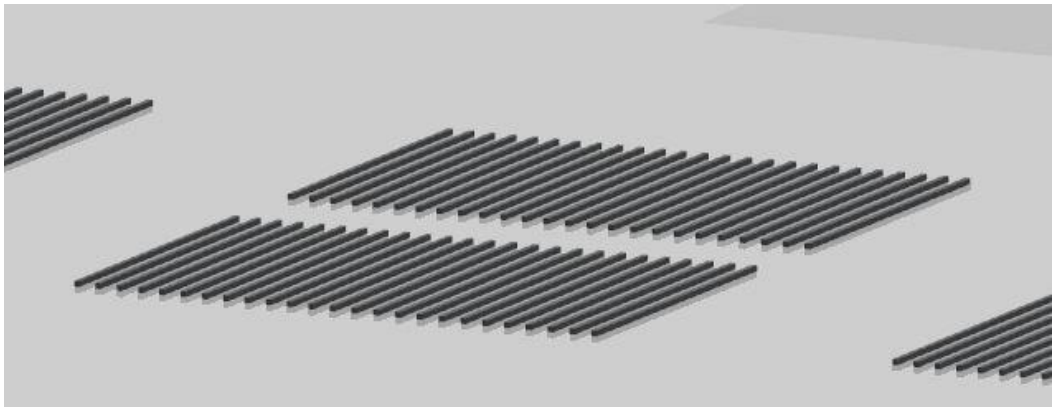


Figure 13 3d view of the design of small channels (first layer) in the mother machine.

2.1.2.1 First layer fabrication using the nanoscribe

1- The fused silica substrate was coated with surpass 3000 by pouring it in a beaker and immersing the substrate in the beaker for 1 minute.

2- The substrate was then taken out and blow dried using N_2.

3- Used Nano 3d printer with a 63x objective and IP-Dip resist. Marked the top part of the substrate for easier alignment. Mounted the substrate onto the holder and dipped a droplet of resist to the center of the substrate. Inserted the sample holder into the stage and printed the pattern. Unloaded the substrate.

4- The printed structure was then developed in SU8 developer for 8 minutes and left to dry at room temperature. The developed layer was then stored overnight at room temperature and exposed to sunlight to allow the binding to stabilize.

2.1.2.2 Second layer fabrication procedure using maskless aligner

1- Coated the substrate using an HMDS oven (YES 3TA HMDS Vapor Prime Oven).

2- Spread the photoresist SU8- 2015 using the spin coater in two stages: 10s at 500 rpm and acceleration 100 $\frac{rpm}{s^2}$, followed by a 60s spin at 1500 rpm and acceleration 300 $\frac{rpm}{s^2}$.

3- Pre- exposure baked the coated substrateat 65℃ for 5 minutes, and then at 95℃ for 50 minutes using a hotplate.

4- Exposed the coated substrate using the maskless aligner with LED light (390 nm - 10000 mW).

5- Post-exposure baked the substrate at 65℃ for 3 minutes, and then at 95℃ for 15 minutes.

6- Developed the exposed photoresist in SU8 developer for 1:45 minutes. The timing of this development was very critical. Extending this time by even few seconds caused the channels to peel off, while shortening it by few seconds resulted in underdeveloped photoresist.
7- Hard baked the substrate at 200 ℃ for 20 minutes.

2.1.3 Silanization

After the required pattern on the wafer (or the fused silica) was achieved, I coated the resulting mold with a silane monolayer. After this coating, no changes could be made to the pattern. This coating ensured that the surface of the mold does not bind to the PDMS, and that PDMS can be peeled off easily. The silanization was achieved as follows:

1- Poured a small amount (~1 ml) of trichlorosilane 99% (Gelest Inc., USA) in a plastic cup.
2- Left the cup and the wafer in vacuum for 30 minutes.

The wafer was then ready for preparing the final microfluidic device with PDMS.

2.2 PDMS fabrication

1- Mixed a 9:1 mass ratio of PDMS oligomers and a cross-linking agent thoroughly.
2- The wafer (or the fused silica) with the desired mold that was created using the methods described in section 2.1, was placed in a plastic petri dish (100×150 mm for silicon wafer and 35×15 mm for fused silica) and the PDMS mixture was poured on the mold.

3- Degassed the PDMS mixture, by incubating it in vacuum for ~20 minutes, or until all bubbles have disappeared from the PDMS mix.
4- Cured the PDMS by incubating the petri dish in an oven overnight at 65℃.
5- The following day, I cut out a small section of the PDMS around the pattern and peeled off the mold. I also cut out additional pieces of PDMS with no patterns to be used for forming the inlet and outlet of the device.
6- Treated all PDMS parts with oxygen plasma for 30 seconds and attached them together.
7- Made two small holes in the inlet and outlet of the machine using a syringe.
8- The PDMS piece together with a microscope cover slip (24×60mm, 0.16 to 0.19 mm thick) were treated with oxygen plasma for 30 seconds to form the complete device (Figure 14).
9- The complete device was then left on a hotplate at 70℃ for one hour to strengthen the binding.

The device was then ready for carrying out the desired experiment.

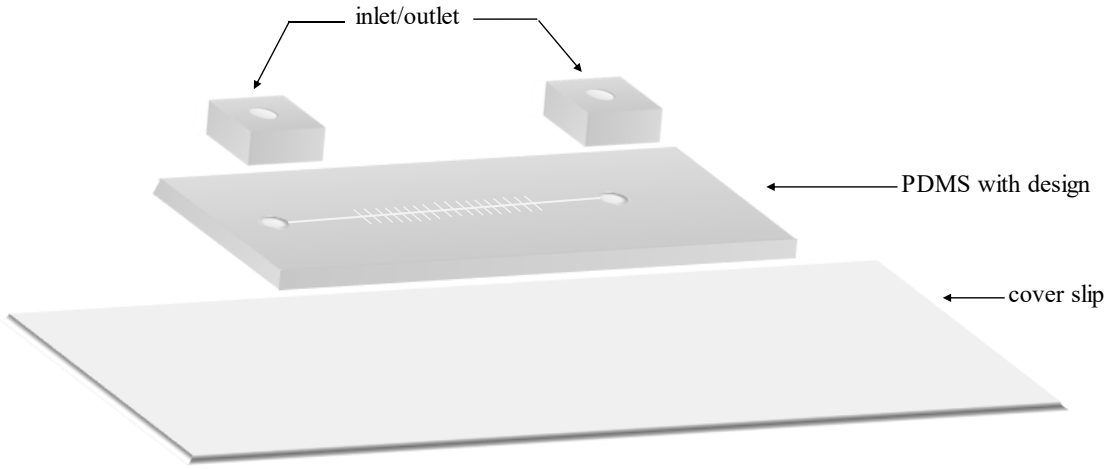


Figure 14 Different parts of the mother machine device. Darker grey is PDMS, and lighter grey is the microscope cover glass. Two small pieces of PDMS are first attached to the main PDMS with the mothermachine design, then they are all attached to the cover slip.

3.0 Correlation among measured properties during cellular growth[68]

Living cells maintain cellular homeostasis despite significant molecular noise affecting all cellular processes[10,12,69–72]. Understanding how cells conserve their properties in the presence of such noise has been a central question in biology for decades. One area of research that falls within the scope of this question, and which has received significant consideration in the last decade, is bacterial cell-size homeostasis[10,12,72–74]. Mother machine experiments have provided high quality cell-size measurements that allow us to examine cell growth and division dynamics quantitatively (see Chapter 2.0), and develop new theoretical models of how cells' growth dynamics contribute to size homeostasis[10,40,64]. Indeed, previous studies have shown strong correlations between particular pairs of measured cellular properties, which inspired various phenomenological models of cell-size control mechanisms[10,64,73,75,76]. However, such correlations exist simultaneously among many cellular properties, which suggests that the dynamics of cell growth and division is determined by multiple cellular mechanisms whose integrated actions give rise to the homeostasis[19,20]. On the other hand, the observed correlations may be indirect and do not necessarily imply that those cellular properties directly depend on one another. To separate indirect correlations from key dependencies between the various measured cellular properties, a systematic approach is required to infer simplified growth mechanism models from high-throughput data.

In this chapter, I present our analysis of *E. coli* single-cell growth and division data obtained by the mother machine, which reveal dependencies that were previously not considered as part of the cell-size homeostasis mechanism. By measuring growth properties, such as growth rate, birth size, and cycle time, of several hundred cells and their descendants, we ask whether these properties of a mother cell can affect the growth of its offspring after several divisions? We

also attempt to find such relationships among properties of sisters and cousin cells. Our results allow for a better description of the growth of daughter cells, sister cells, and cousin cells.

Our analysis is based on applying regression methods to multiple simultaneously measured cellular properties and is not biased by preexisting models or proposed biological mechanisms. We measure cellular properties of sister cells and genealogically more distant cells, as shown in Figure 15, by methods explained in section 3.1. Our model predicts relations between these distant cells that can be tested in the future by designing new experiments.

3.1 Experimental methods

Measuring cell size dynamics was carried out using the mother machine[40,77]. A full explanation of this machine along with how to fabricate it is presented in chapter 2.0.

All measurements were performed using the wild type MG1655 *E. coli* bacteria containing the medium copy-number plasmid pZA[78], expressing Green Fluorescent Protein (GFP) under the control of the viral λ-phage Pr promoter. The bacteria were initially cultured overnight in LB medium at 30°C, diluted in the same medium the following day, and regrown to early exponential phase, Optical Density (OD) between 0.1 and 0.2. When the cells reached the desired OD, they were concentrated into fresh medium to an OD~0.3 and loaded into the microfluidic trapping device. The cells were grown in the traps for tens of generations under constant conditions. The medium (LB) was pumped through the device at a rate of 1 ml/hr.

Images of the channels were acquired every 3 minutes in DIC and fluorescence mode (GFP) using a Hamamatsu ORCA-flash 4.0 camera, mounted on a Nikon Eclipse T*i*2 inverted microscope with a 100X objective. Temperature was stabilized at 32℃ using an Okolab

microscope enclosure. TLMTracker[79] and made-in-house MATLAB codes were used to measure the length of the cells as a function of time. In addition to tracking the "mother" cell trapped at the end of the channel as in previous studies, its immediate sister was also tracked during every generation (Figure 15). These data were then used for further analysis as detailed below.

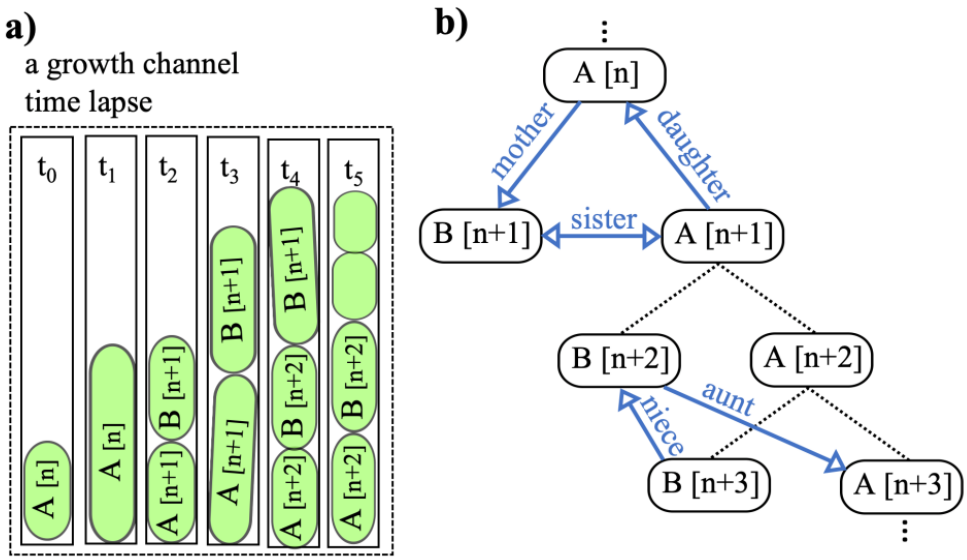


Figure 15 a) A schematic depiction of a single microchannel, and the growth dynamics of trapped cells over time, where *t* labels representative time points, *n* labels the generation number, and A and B denote the cell at the bottom of the channel and its sister, respectively. A typical genealogical tree of the cell lineage is presented in (b), and the relations among different cells are elaborated by the arrows.

3.2 Cell cycle variables are correlated across generations

To begin our analysis, we need to extract dynamical variables that adequately describe the cell growth cycles and cell divisions. To this end, each cell cycle of a temporal sequence of cell length is fitted to an exponential curve:

$$x_n(t) = x_n(0)e^{\alpha_n t} \quad \text{for} \quad 0 \le t \le T_n$$

where the subscript n labels then n -th cell cycle, $x_n(0)$ is the length of the cell at the start of the cell cycle, α_n is the exponential growth rate, and T_n is the cell cycle duration.

Our raw data from image analysis include time points (every 3 min) and lengths of both cells (labeled A and B, where A is the so-called mother cell), as illustrated with blue dots in Figure 16. These data were first parsed into individual cell cycles, labeled by $a = A, B$ and $n = 1, \cdots, N$, where N is the number of cell cycles (generations) in a measured cell lineage. Thus, we have data points in the form $\left(T_{n,i}^a, X_{n,i}^a\right)$, where T represents time, X represents length, $i = 1, \cdots, l_n^a$, where l_n^a is the number of data points within the n-th generation of cell a.

Next, we impose a constraint on our data that requires the sum of the initial sizes of both cells A and B to be equal to the final size of cell A in the previous cell cycle. To ensure all pairs of sisters are measured similarly after a mother's division, we set a threshold for the line between two sisters in our image analysis. This constraint ensures that, when we calculate the fractional size of each cell right after division, the fractions sum to 1. To apply this constraint, which links consecutive cell cycles, we fit the growth curves of all cell cycles simultaneously, instead of fitting each cell cycle independently. The curve fitting is explained in detail in Appendix A. Thus, each cell cycle can be characterized by three variables, which we choose to define as:

$$\chi_n = \log x_n(0),\ \alpha_n, \text{ and } \phi_n = \alpha_n\, \tau_n$$

In a plot of cell length (in log scale) versus time, where a cell cycle is approximately represented by a slanted line segment, these variables represent the initial height, slope, and vertical increment of the line segment (Figure 16).

We represent these variables by a vector $(\chi_n, \alpha_n, \phi_n)^\top$, where $\top$ denotes matrix transpose. Note that the variables are arranged in a causally consistent temporal order, such that in each cell cycle, α_n may depend on χ_n, and ϕ_n may depend on both χ_n and α_n. In addition, all variables in

one cell cycle may in principle depend on their values in previous cell cycles. To complete the link between consecutive cell cycles, we define another variable that represents the fraction in log scaled by which a cell divides, $\psi_n = \log(x_n(0)/x_{n-1}(\tau_{n-1})) = \chi_n - \chi_{n-1} - \phi_{n-1}$. In what follows, we analyze the multivariate time series of $(\chi_n, \alpha_n, \phi_n)$.

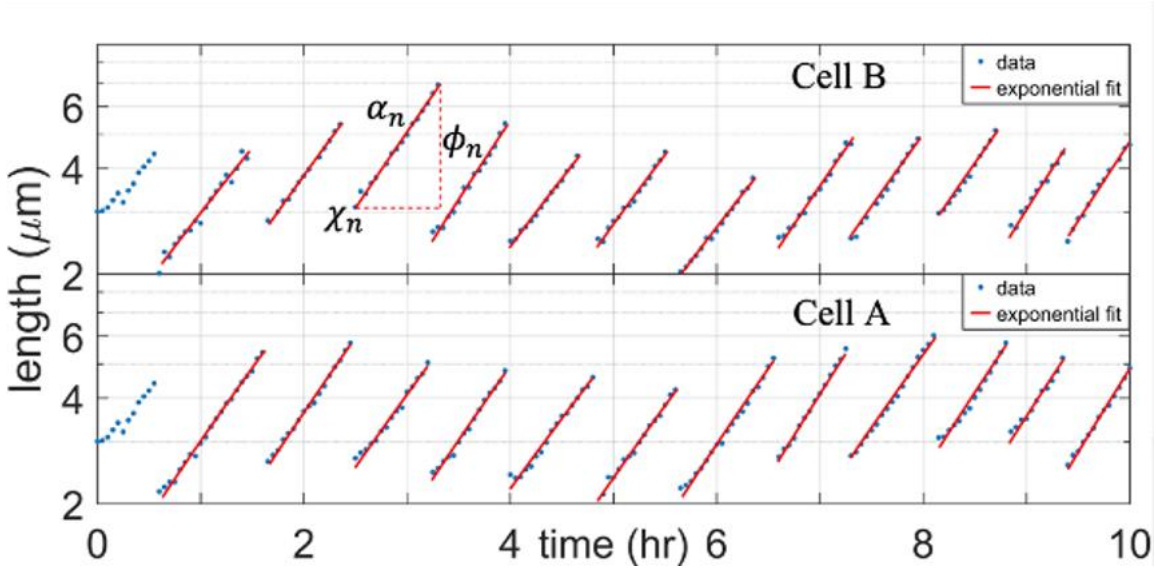


Figure 16 Example traces acquired from the mother machine for two sister cells, where cell A is always the one at the end of the channel and cell B is the sister. Note that at a single time point there can be zero, one, or two B cells being tracked (t1, t2, t4 in Figure 15a, respectively). The red lines are exponential fits (note that the y-axis is in logarithmic scale) calculated using a constraint that ensures the sum of initial sizes of cell A and B from generation n is equal to the final size of cell A from generation n-1 (See Appendix A). See also Appendix Figure 3.

We first examine the temporal correlation between the variables. Black lines in Figure 17 show the correlations between the cell cycle variables as a function of the time delay k in numbers of generations. It can be seen that, for some pairs of variables, such as (ϕ, χ), there is significant correlation over several generations. This suggests that information from one cell cycle is carried over to subsequent cycles, which influences the growth of descendant cells. Moreover, the negative correlation between the pair (ϕ, χ) (Figure 22C and D) implies the existence of a stabilizing

mechanism that controls cell size accumulation and determines the distribution of cell size[10,12,64]. However, correlation does not necessarily imply direct dependence between the variables – two variables may be correlated simply because they both depend on a third variable, which causes indirect correlation. Below we investigate direct dependencies, which would better reveal mechanisms of cell size control.

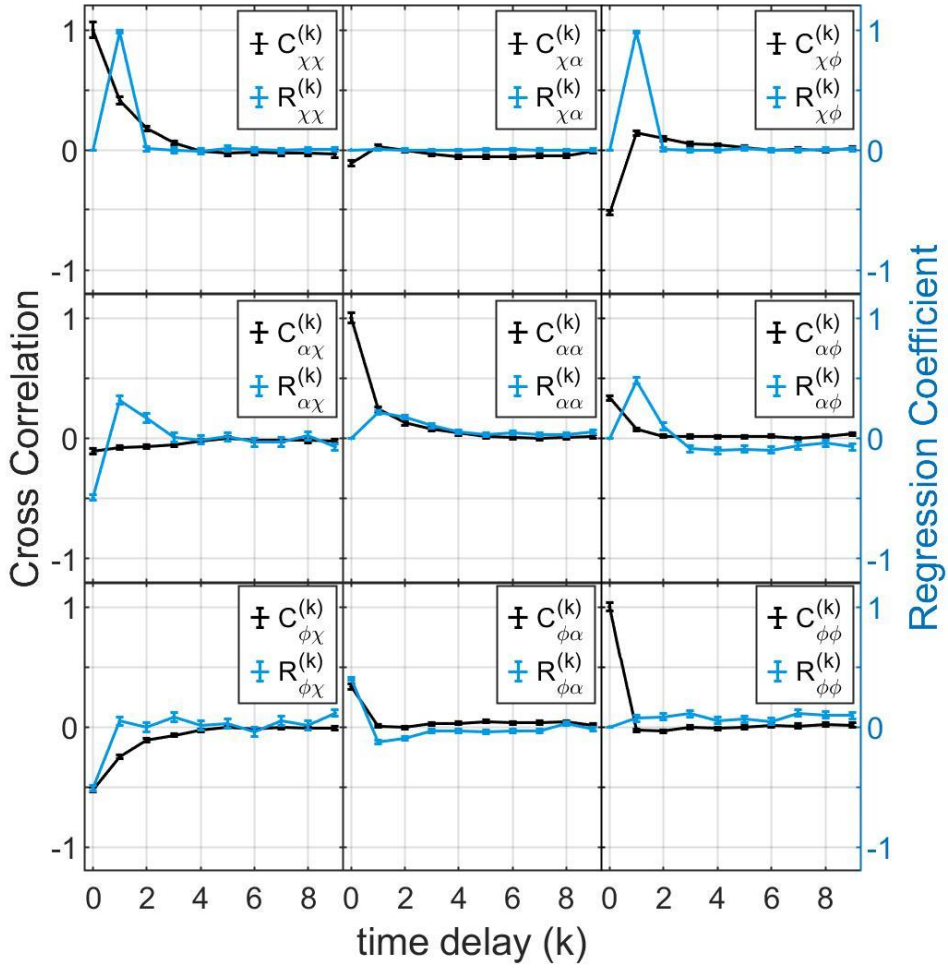


Figure 17 Temporal correlation and regression coefficients between cell cycle variables. Black lines: correlation among three variables (χ_n, α_n, ϕ_n) as a function of the time delay k. Blue lines: regression coefficients calculated using constrained multivariate regression. The coefficients of time delay ≥ 2 are largely negligible; hence, each cell cycle variable depends on the variables of the present and the previous cycle. All error bars are standard errors from 119 lineages (3994 cell cycles).

3.3 Dynamics of cell cycle variables is approximately Markovian

As in regression analysis[80], we fit our time series data to a linear equation:

$$\begin{pmatrix}\chi_n\\ \alpha_n\\ \phi_n\end{pmatrix} = R^{(0)}\begin{pmatrix}\chi_n\\ \alpha_n\\ \phi_n\end{pmatrix} + R^{(1)}\begin{pmatrix}\chi_{n-1}\\ \alpha_{n-1}\\ \phi_{n-1}\end{pmatrix} + \cdots + R^{(p)}\begin{pmatrix}\chi_{n-p}\\ \alpha_{n-p}\\ \phi_{n-p}\end{pmatrix} + C + N_n \qquad \text{Equation 3-1}$$

where $R^{(k)}$ for $k = 0, 1, \cdots, p$ are 3×3 matrices of regression coefficients at time delay k, $C = \left(C^{\chi}, C^{\alpha}, C^{\phi}\right)^{\mathsf{T}}$ is a constant vector, and $N_n = \left(N_n^{\chi}, N_n^{\alpha}, N_n^{\phi}\right)^{\mathsf{T}}$ is the residual (a zero-mean random vector) that represents stochastic variation in cell size dynamics. Note that, due to causal constraints, the matrix $R^{(0)}$ must be strictly lower triangular. This equation describes a multivariate autoregressive process, which we use as an effective model for the dynamics of the variables. The coefficients $R^{(k)}$ can be estimated using our method of "constrained multivariate regression". By simultaneously fitting the coefficients $R^{(k)}$, we essentially remove indirect correlations between the cell cycle variables (χ_n, α_n, and ϕ_n), and reveal direct dependencies among them.

To preserve causality according to the temporal order of variables, the matrix $R^{(0)}$ is constrained to be strictly lower triangular. Thus, the above equation can be written out in component form as:

$$\begin{pmatrix}\chi_n\\ \alpha_n\\ \phi_n\end{pmatrix} = \begin{pmatrix}0 & 0 & 0\\ R_{\alpha\chi}^{(0)} & 0 & 0\\ R_{\phi\chi}^{(0)} & R_{\phi\alpha}^{(0)} & 0\end{pmatrix}\begin{pmatrix}\chi_n\\ \alpha_n\\ \phi_n\end{pmatrix} + \begin{pmatrix}R_{\chi\chi}^{(1)} & R_{\chi\alpha}^{(1)} & R_{\chi\phi}^{(1)}\\ R_{\alpha\chi}^{(1)} & R_{\alpha\alpha}^{(1)} & R_{\alpha\phi}^{(1)}\\ R_{\phi\chi}^{(1)} & R_{\phi\alpha}^{(1)} & R_{\phi\phi}^{(1)}\end{pmatrix}\begin{pmatrix}\chi_{n-1}\\ \alpha_{n-1}\\ \phi_{n-1}\end{pmatrix} + \cdots$$

$$+ \begin{pmatrix}R_{\chi\chi}^{(p)} & R_{\chi\alpha}^{(p)} & R_{\chi\phi}^{(p)}\\ R_{\alpha\chi}^{(p)} & R_{\alpha\alpha}^{(p)} & R_{\alpha\phi}^{(p)}\\ R_{\phi\chi}^{(p)} & R_{\phi\alpha}^{(p)} & R_{\phi\phi}^{(p)}\end{pmatrix}\begin{pmatrix}\chi_{n-p}\\ \alpha_{n-p}\\ \phi_{n-p}\end{pmatrix} + \begin{pmatrix}C^{\chi}\\ C^{\alpha}\\ C^{\phi}\end{pmatrix} + \begin{pmatrix}N_n^{\chi}\\ N_n^{\alpha}\\ N_n^{\phi}\end{pmatrix}$$

To estimate the values of the coefficients $R_{ij}^{(k)}$, we use the least squares method row by row (each row corresponds to a "multiple regression" problem). As an example, for the first row, we can define a vector $X_n = \left(\chi_{n-1}, \alpha_{n-1}, \phi_{n-1}, \cdots, \chi_{n-p}, \alpha_{n-p}, \phi_{n-p}, 1\right)^\top$, so that the equation becomes $\chi_n = \left(R_{\chi\chi}^{(1)}, R_{\chi\alpha}^{(1)}, R_{\chi\phi}^{(1)}, \cdots, R_{\chi\chi}^{(p)}, R_{\chi\alpha}^{(p)}, R_{\chi\phi}^{(p)}, C^\chi\right) X_n + N_n^\chi$. Then the coefficients can be estimated using the standard formula:

$$\left(R_{\chi\chi}^{(1)}, R_{\chi\alpha}^{(1)}, R_{\chi\phi}^{(1)}, \cdots, R_{\chi\chi}^{(p)}, R_{\chi\alpha}^{(p)}, R_{\chi\phi}^{(p)}, C^\chi\right) = \langle \chi_n X_n^\top \rangle \langle X_n X_n^\top \rangle^{-1}$$

where $\langle X_n X_n^\top \rangle$ is the (uncentered) covariance matrix of X_n and $\langle \chi_n X_n^\top \rangle$ is the (uncentered) cross-covariance between χ_n and X_n. Multiplying the cross-covariance by the inverse of the covariance matrix effectively decorrelates the variables in X_n. Moreover, the variance of N_n^χ can be estimated as:

$$\mathbb{V}\left(N_n^\chi\right) = \langle \chi_n \chi_n \rangle - \langle \chi_n X_n^\top \rangle \langle X_n X_n^\top \rangle^{-1} \langle X_n \chi_n \rangle,$$

where $\langle \cdots \rangle$ represent the ensemble average. Finally, the standard errors of the estimated coefficients are:

$$\left(\frac{\mathrm{diag}(\langle X_n X_n^\top \rangle^{-1})}{M} \mathbb{V}\left(N_n^\chi\right)\right)^{1/2},$$

where $\mathrm{diag}(\langle X_n X_n^\top \rangle^{-1})$ represent a vector, whose components are the diagonal elements of the inverse of the matrix $\langle X_n X_n^\top \rangle$, M is the number of data points, and the square root is taken elementwise. The other two rows of coefficients can be estimated in the same way.

Our method of parameter estimation is similar but different from both "multiple regression" and "multivariate regression" often used in statistical analysis[81] (see Table 2). Multiple regression typically deals with one response variable, y, and multiple explanatory variables, x_i's, such that $y = a_1 x_1 + a_2 x_2 + \cdots + c$, where a_i's are the regression coefficients and c is a constant.

If we represent the explanatory variables by a vector $X = (x_1, x_2, \cdots)^\top$ and the regression coefficients by a vector $A = (a_1, a_2, \cdots)$, then $y = A^\top X + c$. On the other hand, multivariate regression deals simultaneously with multiple response variables, y_i's, and multiple explanatory variables, x_i's, such that each response variable y_i can be described as $y_i = a_{i1}x_1 + a_{i2}x_2 + \cdots + c_i$. In this case, using vectors $Y = (y_1, y_2, \cdots)^\top$, $X = (x_1, x_2, \cdots)^\top$ and $C = (c_1, c_2, \cdots)^\top$, the regression coefficients then form a matrix $A = (a_{ij})$ such that $Y = A\,X + C$, where all entries a_{ij} are to be estimated. In the method we used here, we have multiple response variables, $(\chi_n, \alpha_n, \phi_n)$, as well as many explanatory variables, such as $(\alpha_n, \phi_n, \chi_{n-1}, \alpha_{n-1}, \phi_{n-1}, \ldots)$. However, due to constraints stemming from causality, not all entries of the regression coefficient matrix are free parameters. Therefore, we cannot directly apply formulas from multivariate regression that calculate the coefficient matrix as a whole; instead, we treat each row of the matrix separately, removing the non-free parameters and fitting the remaining entries using multiple regression.

The estimated values of the coefficients $R^{(k)}$ are shown in Figure 17 (blue lines) and presented also in Table 3, and the algorithms for calculating them is provided in Appendix B. For details on how the values for the black lines were calculated see section 3.8. The temporal behavior of the coefficients $R^{(k)}$ can be contrasted with that of the correlation functions (black lines in Figure 17) that usually change slowly over a time interval. For example, the regression coefficient $R^{(k)}_{\phi\chi}$ is large and negative for $k = 0$, but quickly vanishes at $k \geq 1$, indicating that much of the correlation between the pair (ϕ, χ) is indirect. In fact, most regression coefficients with time delay $k > 1$ are small and do not significantly affect the correlation functions (Figure 18). This means that the values of the cell cycle variables mainly depend on their values in the previous cell cycle, but not on earlier ones. In other words, the dynamics of these variables is approximately Markovian. We will focus on those regression coefficients with $k \leq 1$.

Table 2 Comparison of "constrained multivariate regression" with other common types of regression methods.

Method	Number of variables	Expected relationship between variables	Parameters to be estimated
simple regression	one response variable (y), one explanatory variable (x)	$y = a\,x + c$	regression coefficient (a), constant (c)
multiple regression (multivariable regression)	one response variable (y), multiple explanatory variables (x_i's, represented by a vector $X = (x_1, x_2, \cdots)^\top$)	$y = \sum_i a_i\,x_i + c$ $= A\,X + c$	regression coefficients (a_i's, represented by a vector $A = (a_1, a_2, \cdots)$), constant (c)
multivariate regression	multiple response variables (y_i's, represented by a vector $Y = (y_1, y_2, \cdots)^\top$), multiple explanatory variables (x_i's, represented by a vector $X = (x_1, x_2, \cdots)^\top$)	$y_i = \sum_j a_{ij}\,x_j + c_i$ $Y = \boldsymbol{A}\,X + C$	regression coefficients (a_{ij}'s, represented by a matrix $\boldsymbol{A} = \begin{pmatrix} a_{11} & a_{12} & \cdots \\ a_{21} & a_{22} & \cdots \\ \vdots & \vdots & \ddots \end{pmatrix}$), constants ($c_i$'s, represented by a vector $C = (c_1, c_2, \cdots)^\top$)
constrained multivariate regression	multiple response variables (y_i's, represented by a vector $Y = (y_1, y_2, \cdots)^\top$), multiple explanatory variables (x_i's, represented by a vector $X = (x_1, x_2, \cdots)^\top$)	$y_i = \sum_j a_{ij}\,x_j + c_i$ $Y = \boldsymbol{A}\,X + C$	like above, but where certain entries of the regression coefficient matrix are constrained to take prefixed values. For example, the coefficient matrix $R^{(0)}$ in Eq. (1) is causally constrained to be strictly lower triangular. In such a case we must have $\boldsymbol{A} = \begin{pmatrix} 0 & 0 & 0 & \cdots \\ a_{21} & 0 & 0 & \cdots \\ a_{31} & a_{32} & 0 & \cdots \\ \vdots & \vdots & \ddots & \ddots \end{pmatrix}$

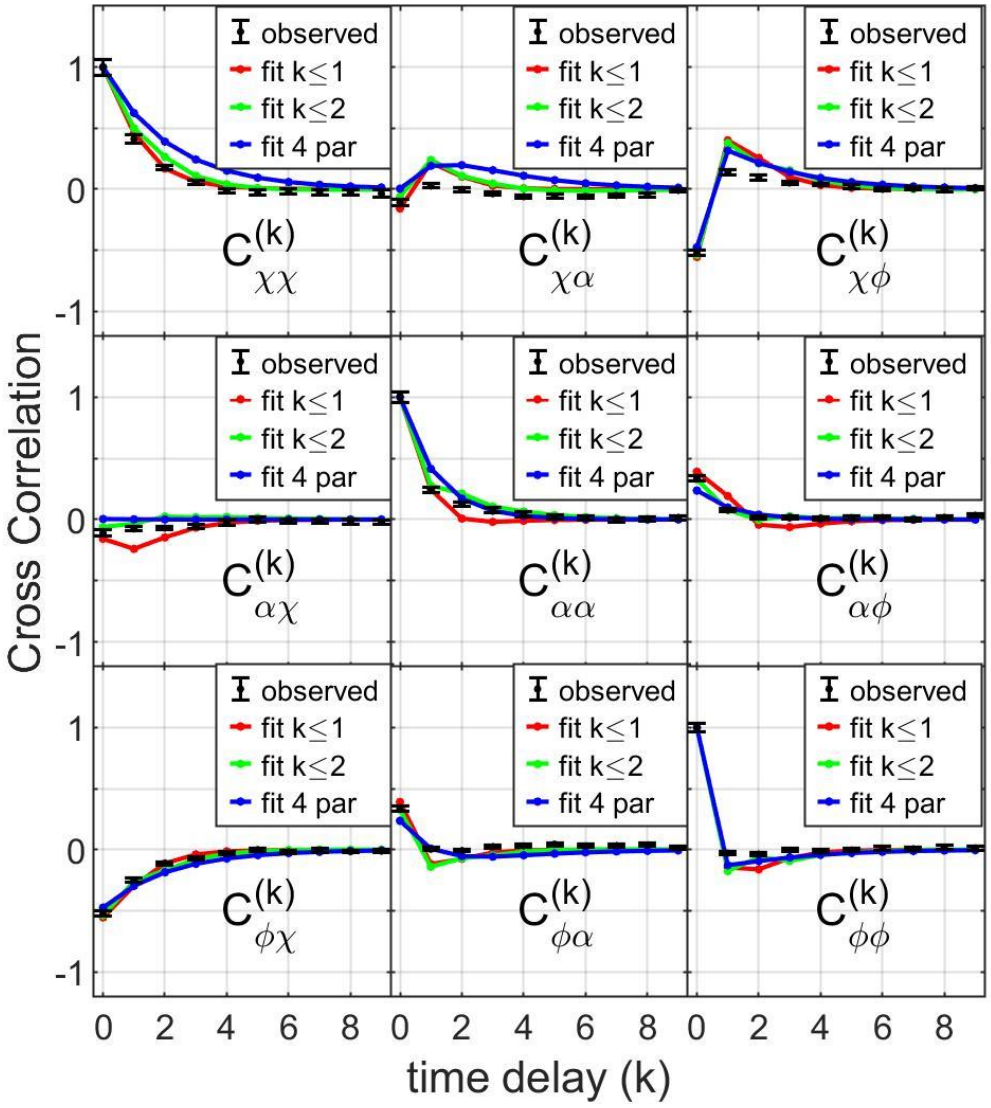


Figure 18 Cross-correlation between cell cycle variables calculated from experimental data and from the estimated parameters of our model. Black dots: cross-correlation observed in the experimental data; red line: cross-correlation calculated using all regression coefficients $R^{(k)}_{ij}$ with $k \leq 1$; green line: cross-correlation calculated using all regression coefficients $R^{(k)}_{ij}$ with $k \leq 2$; blue line: cross-correlation calculated using only 4 parameters from Table 4.

Table 3 Values of the first nine regression coefficients, constants, and variances estimated using our multi-regression analysis (see Figure 17).

regression coefficient	matrix elements		
$R^{(0)}$	0	0	0
	–0.49 ± 0.02	0	0
	–0.51 ± 0.02	0.40 ± 0.01	0
$R^{(1)}$	0.98 ± 0.01	–0.00 ± 0.01	0.98 ± 0.01
	0.32 ± 0.03	0.22 ± 0.02	0.48 ± 0.03
	0.05 ± 0.04	–0.12 ± 0.02	0.07 ± 0.03
$R^{(2)}$	0.01 ± 0.02	0.00 ± 0.01	0.00 ± 0.02
	0.17 ± 0.04	0.18 ± 0.02	0.10 ± 0.03
	0.00 ± 0.04	–0.10 ± 0.02	0.08 ± 0.03
$R^{(3)}$	0.00 ± 0.02	0.00 ± 0.01	0.00 ± 0.02
	0.01 ± 0.04	0.10 ± 0.02	-0.09 ± 0.03
	0.08 ± 0.04	-0.04 ± 0.02	0.11 ± 0.03
$R^{(4)}$	-0.01 ± 0.02	-0.01 ± 0.01	0.00 ± 0.02
	-0.01 ± 0.04	0.05 ± 0.02	-0.10 ± 0.03
	0.01 ± 0.04	-0.04 ± 0.02	0.05 ± 0.03
$R^{(5)}$	0.01 ± 0.02	0.00 ± 0.01	0.01 ± 0.02
	0.01 ± 0.04	0.03 ± 0.02	-0.09 ± 0.03
	0.03 ± 0.04	-0.04 ± 0.02	0.06 ± 0.03
$R^{(6)}$	0.00 ± 0.02	0.00 ± 0.01	0.00 ± 0.02
	-0.03 ± 0.04	0.05 ± 0.02	-0.10 ± 0.03
	-0.04 ± 0.04	-0.03 ± 0.02	0.05 ± 0.03
$R^{(7)}$	-0.01 ± 0.02	0.00 ± 0.01	0.00 ± 0.02
	-0.03 ± 0.04	0.03 ± 0.02	-0.06 ± 0.03
	0.05 ± 0.04	-0.04 ± 0.02	0.12 ± 0.03
$R^{(8)}$	0.00 ± 0.02	0.00 ± 0.01	0.00 ± 0.02
	0.02 ± 0.04	0.03 ± 0.02	-0.04 ± 0.03
	0.01 ± 0.04	0.03 ± 0.02	0.10 ± 0.03
$R^{(9)}$	0.00 ± 0.01	0.00 ± 0.01	0.00 ± 0.01
	-0.07 ± 0.03	0.05 ± 0.01	-0.07 ± 0.03
	0.11 ± 0.03	-0.02 ± 0.01	0.09 ± 0.03
C	–0.67 ± 0.03		
	0.32 ± 0.05		
	0.38 ± 0.05		
$\mathrm{diag}(\Sigma)^{1/2}$	0.06		
	0.11		
	0.12		

3.4 Cell size dynamics can be described by just a few parameters

Among the 18 entries of the matrices of regression coefficients $R^{(0)}$ and $R^{(1)}$, there are in principle 12 free parameters. This is because $R^{(0)}$ is constrained by causality to be strictly lower triangular as explained in the previous section, which eliminates 6 parameters. Furthermore, because cell divisions appear to be symmetric on average, the 3 entries in the top row of $R^{(1)}$ are expected to be $(1, 0, 1)$, which is confirmed by our regression results. To prove this, recall that e^{ψ_A} and e^{ψ_B} are the size fractions that sister cells, A and B, receive from their mother, which should sum to 1. Since the division asymmetry is small, we can write $\psi_a = \log\frac{1}{2} + \Delta\psi_a$ for $a = A, B$ and write the sum as: $e^{\psi_A} + e^{\psi_B} = \frac{1}{2}e^{\Delta\psi_A} + \frac{1}{2}e^{\Delta\psi_B} = 1$. Expanding to first order in $\Delta\psi_a$ yields: $\Delta\psi_A + \Delta\psi_B = 0$, and thus $\psi_A + \psi_B = 2\log\frac{1}{2}$. We first use this relation to derive the constraints on the regression coefficients. Since by definition, $\psi_a = \chi_a - \chi_M - \phi_M$ for $a = A, B$, we expect:

$$\chi_A + \chi_B = 2\chi_M + 2\phi_M + \psi_A + \psi_B = 2\chi_M + 2\phi_M + 2\log\frac{1}{2}$$

On the other hand, our model gives $\chi_a = R_{11}^{(1)}\chi_M + R_{12}^{(1)}\alpha_M + R_{13}^{(1)}\phi_M + C^{\chi} + N_a^{\chi}$ for $a = A, B$. Thus, by adding the sister cells, we obtain:

$$\chi_A + \chi_B = 2R_{11}^{(1)}\chi_M + 2R_{12}^{(1)}\alpha_M + 2R_{13}^{(1)}\phi_M + 2C^{\chi} + (N_A^{\chi} + N_B^{\chi})$$

Comparing the coefficients in the above two equations leads to: $\left(R_{11}^{(1)}, R_{12}^{(1)}, R_{13}^{(1)}\right) = (1, 0, 1)$.

Moreover, initial regression results show that the 3 entries in the bottom row of $R^{(1)}$ are approximately zero (Figure 17) and so will be neglected. There are 6 entries of the regression coefficient matrices left with non-zero values: $R_{\alpha\chi}^{(0)}$, $R_{\phi\chi}^{(0)}$, $R_{\phi\alpha}^{(0)}$, $R_{\alpha\chi}^{(1)}$, $R_{\alpha\alpha}^{(1)}$, and $R_{\alpha\phi}^{(1)}$. Due to the

proximity between the absolute values of the entries $R^{(0)}_{\alpha\chi}$, $R^{(1)}_{\alpha\chi}$, and $R^{(1)}_{\alpha\phi}$, we initially parameterized the matrices $R^{(0)}$ and $R^{(1)}$ as:

$$R^{(0)} = \begin{pmatrix} 0 & 0 & 0 \\ -\kappa + \varepsilon & 0 & 0 \\ -\beta & \gamma & 0 \end{pmatrix} \quad \text{and} \quad R^{(1)} = \begin{pmatrix} 1 & 0 & 1 \\ \kappa & \lambda & \kappa + \delta \\ 0 & 0 & 0 \end{pmatrix}$$

The estimated values of those 6 parameters calculated in five different experiments are shown in Figure 19. Among those values, the value of δ estimated from all the experiments is consistently small, with an average 0.01 ± 0.01, which is insignificant; the value of ε has an average of -0.11 ± 0.01, which is small compared to κ and, when estimated using data from different experiments, it fluctuates around zero (Figure 19). We therefore decided to drop these 2 parameters; thus, this eventually brings us down to 4 parameters in total.

We can now write the regression coefficients as:

$$R^{(0)} = \begin{pmatrix} 0 & 0 & 0 \\ -\kappa & 0 & 0 \\ -\beta & \gamma & 0 \end{pmatrix} \quad \text{and} \quad R^{(1)} = \begin{pmatrix} 1 & 0 & 1 \\ \kappa & \lambda & \kappa \\ 0 & 0 & 0 \end{pmatrix}$$

The 4 parameters κ, λ, β, and γ have clear interpretations that will be discussed below. The values of these parameters are re-estimated under the additional constraints that the remaining entries of the matrices are fixed to 0 or 1, and the results are shown in Table 4 and Figure 20. These four parameters effectively summarize the statistical dependencies among the cell cycle variables measured in our experiments.

To verify the effectiveness of our model, we simulated cell growth and division using a null model based on the adder mechanism alone which is explained in detail in Appendix C. In that case, all parameters except β should be zero. We then estimated the values of the parameters from the simulated data. These estimates provide a baseline for the systematic errors of the parameters. Using this null model, we verified that the values of the parameters we found from our

experiments are significant compared to their baselines. These baselines are shown in Figure 19 with blue dots.

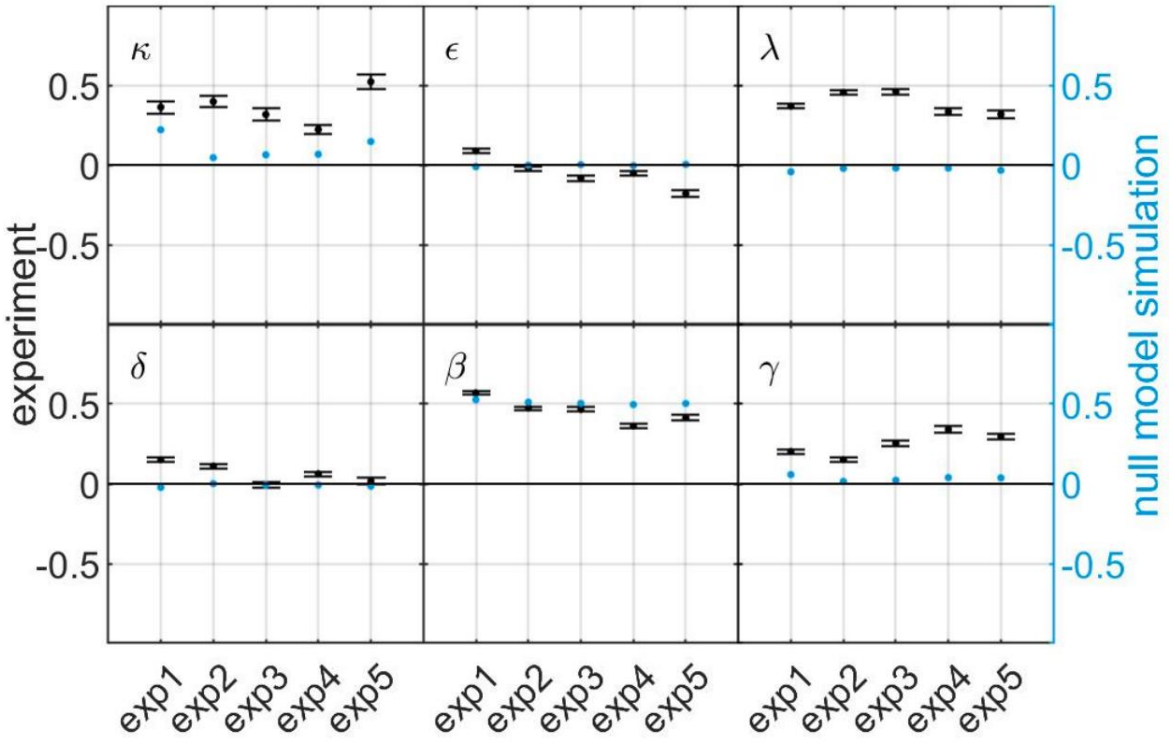


Figure 19 Comparison of estimated parameters across experiments with their baselines. Five separate experiments with the same conditions were performed to confirm the consistency in the values of estimated model parameters. Experiments 1–5 include 1953, 1980, 1363, 1453, and 1178 cell cycles, respectively. The blue dots represent the baselines for the parameters found using the simulated null model (see Appendix C).

Table 4 Estimated parameters using constrained multivariate regression.

	κ	λ	β	γ
Constrained multivariate regression (Figure 18)	0.43 ± 0.02	0.41 ± 0.01	0.38 ± 0.01	0.26 ± 0.01
simple regression (Figure 19)	0.42 ± 0.02	0.41 ± 0.01	0.41 ± 0.01	0.34 ± 0.01

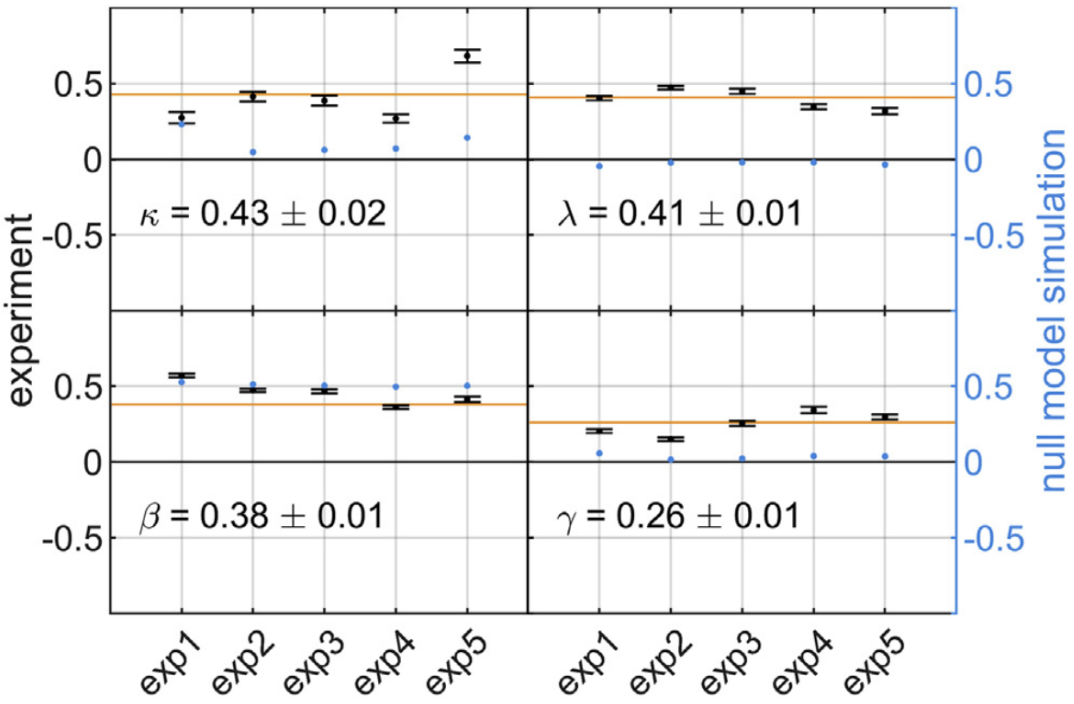


Figure 20 Comparison of estimated parameters across experiments with their baselines. Five separate experiments with same conditions were performed to confirm the consistency in the values of estimated model parameters. Experiments 1–5 include 1953, 1980, 1363, 1453, and 1178 cell cycles, respectively. The blue dots represent the baselines for the parameters found using the simulated null model (see Appendix C). Except for β which was set to 0.5 in the simulation, the other parameters all have a low baseline, validating the significance of the parameters estimated from our experimental data.

3.5 Parameter values vary between experiments

Although the estimated parameter values are significant compared to their baselines in each experiment, their values vary among different experiments (Figure 20). This variation is larger than the population variability measured in all experiments and is not due to the lack of statistical significance. In order to determine if this variation is related to other differences between the experiments, we examined the distribution of various measurable properties of the cells in each

experiment, such as the growth rate, cell-cycle time, and cell size. We found that, indeed, there are slight differences in those cellular properties between the experiments (Figure 21). Note that these differences emerged despite our effort to maintain the experimental conditions the same for all experiments. They may have resulted from variations in some environmental conditions that are beyond the extent of our control. To account for the differences in the distribution of cellular properties, we tried standardizing (removing the mean and normalizing by the standard deviation) each cell cycle variable before fitting the parameters in each experiment. The results still exhibited similar variation between the experiments (Figure 19). This indicates that environmental variations do not only affect the mean and distribution of those variables but also their dependencies, such as the dependency of total size accumulation on birth length (β) or growth rate (γ).

The variation of the parameter values shows that, on the one hand, the current experimental setup is imperfect and can be further improved to reduce environmental effects, as we discuss later. On the other hand, those parameters should be considered as effective descriptions of the cell growth dynamics and may not be universal across environmental conditions. In particular, there may be multiple mechanisms acting in tandem to maintain cellular homeostasis, and their relative contributions to the overall observed behavior of the cell may change in different conditions. Therefore, in analyzing experimental data, we should not be restricted to preexisting models that often focus on one mechanism and are characterized by few predefined parameters. Our regression analysis, which is not biased by preconceived models, provides an "agnostic" way of detecting effective parameters that characterize cell growth and size control.

Despite the quantitative variation of these parameters, the qualitative dependencies between cell cycle variables are robust among experiments (Figure 20). They can be used to reveal possible control mechanisms of cell growth that were previously undetected. In all the analyses

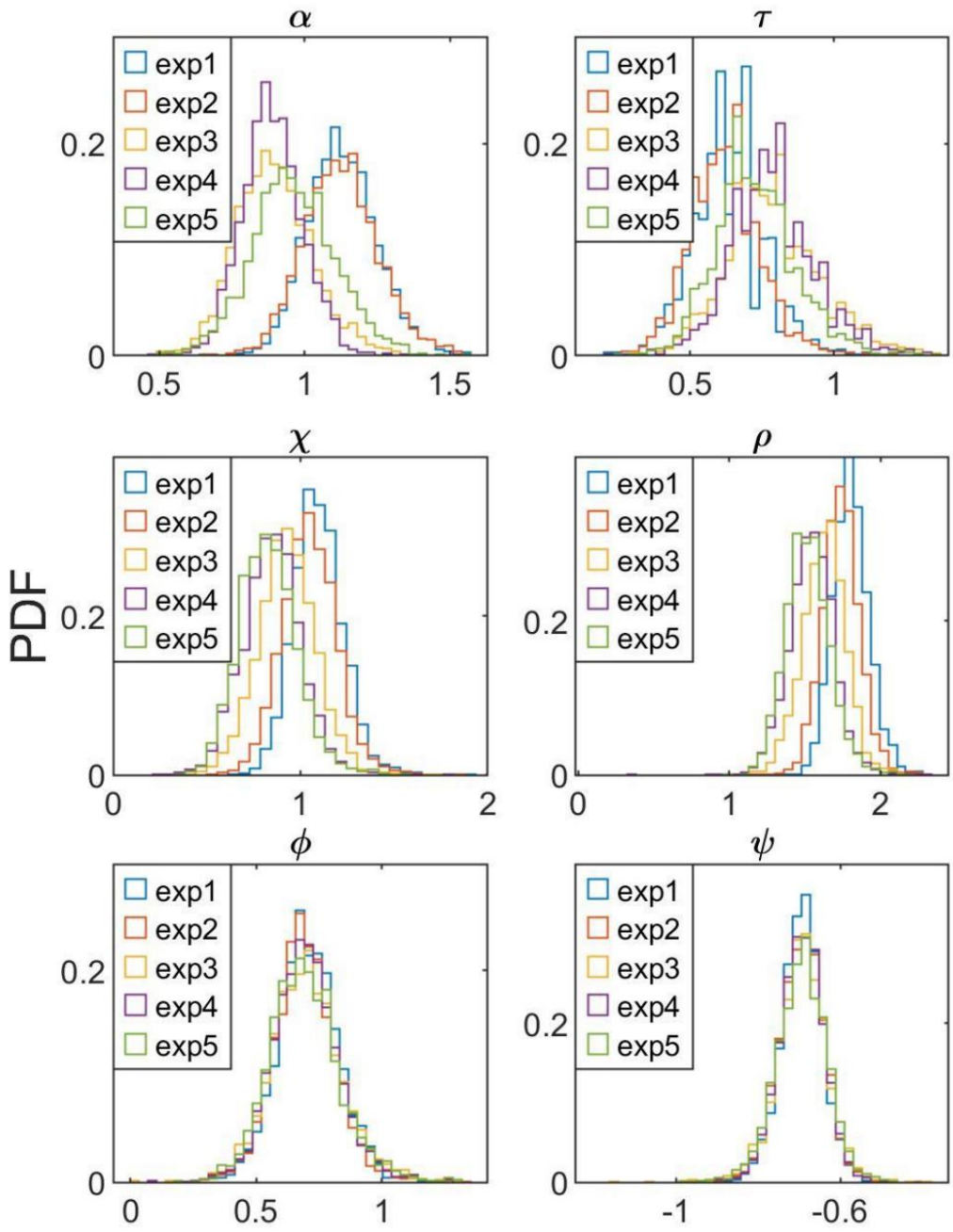


Figure 21 Distribution of measured cellular properties in different experiments. Variables α, χ, ϕ and ψ represent cell growth rate (hr^{-1}), initial size (log scale) (μm), size accumulation (log scale), and division fraction (log scale), respectively, as defined in section 3.2. τ represents cell cycle duration (hr); and $\rho = \chi + \phi$ represents final size (log scale) before division (μm). Data from experiments 3-5 have more similar distributions of the variables and are used for analyses in this chapter.

we present, we combine the data of experiments 3–5 since the distributions of all variables obtained in these experiments are closest to each other (Figure 21). The quantitative values of the parameters we find from these analyses (Table 4) are representative of the dependencies between the cell cycle variables.

3.6 Effective parameters imply mechanisms of cell size regulation

Using the parameters found in section 3.4, we can write the dynamics of Equation 3-1 with time delay $k \leq 1$ as:

$$\chi_n = \chi_{n-1} + \phi_{n-1} + C^{\chi} + N_n^{\chi} \qquad \text{Equation 3-2}$$

$$\alpha_n = -\kappa\, \psi_n + \lambda\, \alpha_{n-1} + C^{\alpha} + N_n^{\alpha} \qquad \text{Equation 3-3}$$

$$\phi_n = -\beta\, \chi_n + \gamma\, \alpha_n + C^{\phi} + N_n^{\phi} \qquad \text{Equation 3-4}$$

Recall that the variables χ_n, α_n, and ϕ_n represent the initial size (in log scale) of a cell, its growth rate during the cell cycle, and the relative amount of growth (in log scale) by the end of the cell cycle (Figure 16). These equations describe previously known features of cellular growth and division dynamics and reveal new details that were never discussed before. Equation 3-2 can be rewritten as $\psi_n = C^{\chi} + N_n^{\chi}$, where $\psi_n = \chi_n - \chi_{n-1} - \phi_{n-1}$ is the fraction of cell size (in log scale) that a daughter cell receives from its mother following the cell division. This equation implies that the size fraction a cell receives from its mother is random and independent of other variables. Such independence on the variables of the mother cell can be understood from the symmetry between two daughter cells. On the one hand, we expect our equation to be applicable to both daughter cells, so their size fractions should depend on the mother cell in the same way.

On the other hand, their size fractions should sum to 1, a constant, which then implies that neither of the fractions can depend on the variables of the mother cell. The symmetry between the daughter cells is supported by our measurements. The regression analysis results in an estimated value of C^{χ} being -0.67 ± 0.03, which implies an average size fraction of 0.51 ± 0.02 (mean $\pm$ SE), consistent with symmetric division. The measured variance of C^{χ} among cells (0.004) means that the size fraction can fluctuate by about 6% from cell to cell.

Equation 3-3 suggests that the growth rate of the cell depends on the growth rate of its mother, and on the size fraction it received from its mother during the division. The parameter λ can be interpreted as a kind of inertia, by which a cell that grew faster in the previous cell cycle tends to grow faster in the current cell cycle. A possible mechanism for such inertia is that the growth rate of the cell depends, in part, on the concentration of key cellular factors that are responsible for the production of various structural and metabolic elements, such as ribosomes, RNA polymerases, and ATPases. Since the concentration of these factors is roughly conserved during divisions, it is expected that the growth rate would persist to some extent in subsequent cell cycles. On the other hand, the parameter κ, following a negative sign, acts as a form of growth compensation, i.e., the smaller daughter cell tends to grow faster and make up for the size difference acquired during the division. Such a negative correlation between the growth rate of sister cells and their relative sizes has also been found in plants[82]. More recently, it was observed that the growth rate of *Bacillus subtilis* is inversely proportional to the cell size at the start of the cell cycle, and changes as the cell cycle advances[17]. Note however, that in our case, it is not the *absolute* size of the daughter, but rather the *relative* size compared to the mother, that affects the growth rate. This is an important distinction, which we discuss in the following section.

Equation 3-4 implies that the relative amount by which a cell grows during each cell cycle is regulated by its initial size and the growth rate during the same cell cycle. Such dependence reflects potential mechanisms of cell size control. The $-\beta$ term represents a negative control by which the cells that are bigger at the beginning of a cell cycle will grow relatively less during that cycle. This control mechanism would be equivalent to the "adder" mechanism (by which a cell adds a fixed amount of size before division) if $\beta = 0.5$ [12,64]. On the other hand, the γ term represents a positive control by which cells that grow faster will have more relative growth during the cell cycle. This would agree with the "timer" mechanism (by which a cell grows for a fixed period of time before division), if γ represents the period of time that the cell is allowed to grow. The fact that both terms are found experimentally to be non-zero suggests that cell size control is complex and might involve more than one mechanism working in tandem, and that both adder and timer are only effective consequences of some unknown actual control mechanism. Indeed, a combination of these mechanisms (a timer phase followed by an adder phase) was proposed to underlie size-homeostasis in the bacteria *Caulobacter crescentus*[72].

3.7 Sister cells comparison confirms growth rate dependence on size fraction

To further verify our regression analysis and the model resulting from it, we conducted further experiments, in which we tracked both daughter cells after each cell division (Figure 15 and Figure 16). One of the cells, the one at the end of the microfluidic trap, commonly known as the "mother" cell, forms a continuous lineage and has been the subject of many previous studies that resulted in several predictions of cell size control mechanisms[10,73,64,83]. The other cell is the

"sister" of the mother cell created after each division (Figure 15.b). From such experiments we extracted the same set of cell cycle variables as before, now for both cells.

With such new types of dataset, we can directly examine the prediction of Equation 3-3. In this equation, the variables for the n-th generation will now represent each of the daughter cells, and the variables for the $(n-1)$-th generation will represent their common mother cell. Accordingly, the growth rates of both daughter cells are:

$$\alpha_A = -\kappa\,\psi_A + \lambda\,\alpha_M + N_A^{\alpha} \quad \text{and} \quad \alpha_B = -\kappa\,\psi_B + \lambda\,\alpha_M + N_B^{\alpha}$$

where the subscripts A and B denote the two sister cells, and M their common mother cell. Adding and subtracting these two equations yield:

$$\Delta\alpha = -\kappa\,(\Delta\psi) + \Delta N^{\alpha} \quad \text{and} \quad \Sigma\alpha = -\kappa\,(\Sigma\psi) + \lambda\,(2\alpha_M) + \Sigma N^{\alpha}$$

where $\Delta\alpha = \alpha_A - \alpha_B$, $\Sigma\alpha = \alpha_A + \alpha_B$, and similarly for other variables. Note that ΔN^{α} and ΣN^{α} are random variables, and $\Sigma\psi \approx 2\log\frac{1}{2}$ (see section 3.4), which is constant. Therefore, simple linear regression of $\Delta\alpha$ over $\Delta\psi$ and of $\Sigma\alpha$ over $2\alpha_M$ should yield the slopes $-\kappa$ and λ, respectively.

The results of such linear regression are shown in Figure 22A and B. The inferred slopes match the estimated values of $-\kappa$ and λ (Figure 20) very well. Similarly, according to Equation 3-4, we did simple linear regression of ϕ_n as a function of χ_n (see also [64]), and α_n respectively (Figure 22C and D), using data from both sister cells. In this case, the inferred slopes show deviations from the values of $-\beta$ and γ obtained using constrained multivariate regression. This is because the latter accounts for the correlation between the variables χ_n and α_n, whereas simple linear regression does not.

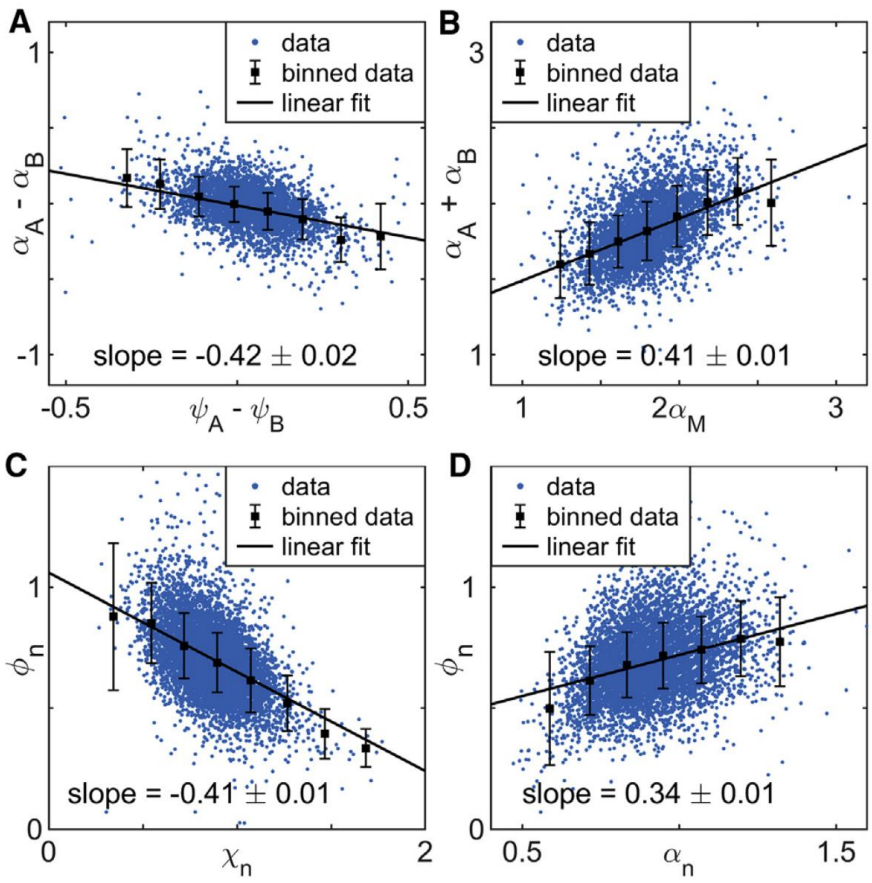


Figure 22 Simple regression between selected variables of both sister cells. (A) The difference between the growth rates (α_n) of two sister cells plotted as a function of the difference between the natural logarithm of their fractions (ψ_n). Subscript A denotes the sister cell trapped at the end of the channel, and B denotes its sister adjacent to it. (B) The sum of the growth rates of sister cells plotted as a function of twice the growth rate of their common mother (M). The first two plots contain N = 3942 pairs of sister cells. (C) The relative amount of growth (ϕ_n) during a single cell cycle plotted as a function of log initial length (χ_n) of the same cycle. (D) ϕ_n plotted as a function of growth rate (α_n) in the same cell cycle. The latter two plots contain N=8173 cells. Black squares and error bars in all plots depict the average and standard deviation of uniformly binned data, respectively. Black lines depict the best linear fit, and the values of the slopes are presented in the plots. The slopes in plots A and B are in good agreement with $-\kappa$ and λ calculated from constrained multivariate regression. The slopes in plots C and D show deviations from $-\beta$ and γ estimated using constrained multivariate regression, because simple regression does not account for the correlation between χ_n and α_n.

3.8 Comparison of sister cells reveals additional correlation in their growth

It is expected that two sister cells sharing the same mother (we can alternatively name this pair as mother and daughter), would exhibit correlations among their cell cycle variables. Such correlations can be calculated from our model, since our description of the dynamics (Equation 3-1 with time delay $k \leq 1$) does not differentiate between the two cells and can be applied to each of them separately.

As shown in Figure 23A, we find that there is a significant difference in the correlations between the sister cells calculated using experimentally observed data (OBS) and predicted by the model (PM). The missing correlation suggests that our model needs to be generalized. The issue is that, like other cell growth models, it only describes the dynamics along a single branch of the lineage tree. Indeed, the model given by Equation 3-1 describes the conditional probability distribution of the cell cycle variables of the daughter cell given those of the mother. However, such a model only captures the probability distribution of *one* daughter cell given the mother (heuristically, $P(A|M)$), instead of the joint distribution of *both* daughters given the mother (i.e., $P(A,B|M)$). In general, the two daughter cells need not be conditionally independent (i.e., $P(A,B|M) \neq P(A|M)\,P(B|M)$). For example, the size fraction that one cell receives from its mother is negatively correlated with the fraction its sister receives. Such conditional dependence would give rise to additional correlation between the sister cells, as observed in the experiments.

We compare the correlation predicted by the model with the correlation calculated directly from the data as follows. We start from the regression coefficients $R^{(0)}$, $R^{(1)}$, the constant C and the diagonal covariance matrix of the residuals Σ, whose diagonal elements are $\left(\mathbb{V}\left(N_n^{\chi}\right), \mathbb{V}(N_n^{\alpha}), \mathbb{V}\left(N_n^{\phi}\right)\right)$. We first calculate the covariance matrix $\Gamma^{(k)}$ as a function of time delay

k. This is done by solving the discrete Lyapunov equation (e.g., by vectorization): $\Gamma^{(0)} = P\,\Gamma^{(0)}\,P^{\top} + Q$, where $P = \left(I - R^{(0)}\right)^{-1} R^{(1)}$ and $Q = \left(I - R^{(0)}\right)^{-1} \Sigma \left(I - R^{(0)}\right)^{-\top}$, and then iteratively calculating: $\Gamma^{(k)} = P\,\Gamma^{(k-1)}$. To calculate the correlations of the cell cycle variables (

Figure 17), we normalize the covariance matrix $\Gamma^{(k)}$ by the variances at $k = 0$, i.e., $C_{ij}^{(k)} = \Gamma_{ij}^{(k)} / \left(\Gamma_{ii}^{(0)}\,\Gamma_{jj}^{(0)}\right)^{1/2}$.

We can write Equation 3-1 with time delay $k \leq 1$ for each cell as

$$\begin{pmatrix} \chi_A \\ \alpha_A \\ \phi_A \end{pmatrix} = R^{(0)} \begin{pmatrix} \chi_A \\ \alpha_A \\ \phi_A \end{pmatrix} + R^{(1)} \begin{pmatrix} \chi_M \\ \alpha_M \\ \phi_M \end{pmatrix} + C + N_A,$$

$$\begin{pmatrix} \chi_B \\ \alpha_B \\ \phi_B \end{pmatrix} = R^{(0)} \begin{pmatrix} \chi_B \\ \alpha_B \\ \phi_B \end{pmatrix} + R^{(1)} \begin{pmatrix} \chi_M \\ \alpha_M \\ \phi_M \end{pmatrix} + C + N_B$$

Equation 3-5

then allow the residuals N_A and N_B to be correlated. The incorporation of the measurement uncertainties are explained in detail in Appendix A. The cross-correlation between N_A and N_B will represent the additional sister-cells correlation. Let the cross-covariance between the residuals N_A and N_B be: $\Sigma' = \langle N_A N_B^{\top} \rangle$. The cross-covariance matrix between the sister cells can be calculated as: $\Gamma'^{(0)} = P\,\Gamma^{(0)}\,P^{\top} + Q'$, where $Q' = \left(I - R^{(0)}\right)^{-1} \Sigma' \left(I - R^{(0)}\right)^{-\top}$. Then the sister-sister correlation $C'^{(0)}$ (Figure 23A) is calculated by: $C_{ij}'^{(0)} = \Gamma'^{(0)}_{ij} / \left(\Gamma_{ii}^{(0)}\,\Gamma_{jj}^{(0)}\right)^{1/2}$.

Furthermore, the cross-covariance matrix between a niece and an aunt cell is given by: $\Gamma'^{(1)} = P\,\Gamma'^{(0)}$, and the niece-aunt correlation $C'^{(1)}$ (Figure 23B) is then calculated as: $C_{ij}'^{(1)} = \Gamma'^{(1)}_{ij} / \left(\Gamma_{ii}^{(0)}\,\Gamma_{jj}^{(0)}\right)^{1/2}$. Similarly, the cross-covariance matrix between cousin cells can be calculated as: $P\,\Gamma'^{(0)}\,P^{\top}$, and their cross-correlation can be obtained by normalization (Figure 24).

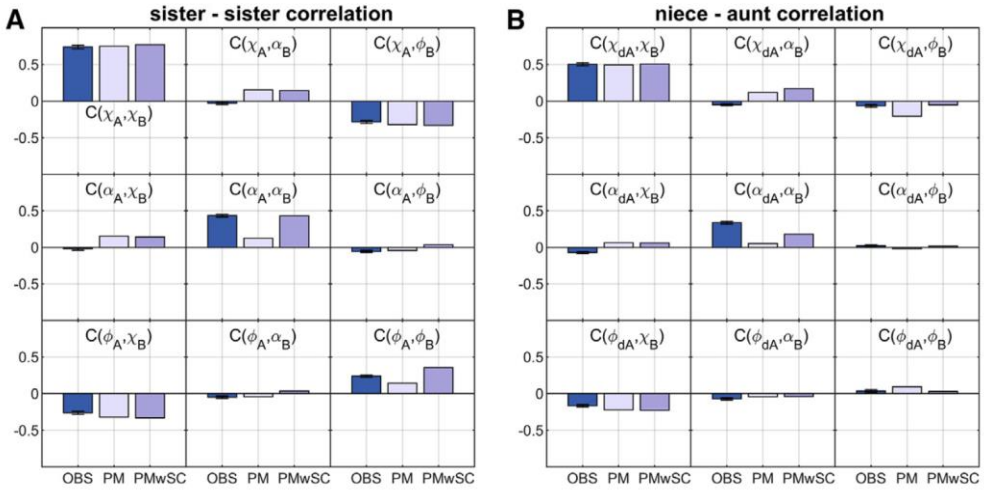


Figure 23 Effect of additional correlation between sister cells on the predictions of our model. Cross-correlations between cell cycle variables of sister A and sister B are calculated using three different methods: the observed values calculated from experimental data (OBS), the predicted values calculated without incorporating additional correlation between sister cells (PM), and the new values calculated after adding the correlation between sister cells (PMwSC). (A) Cross-correlations between sister A and sister B. (B) Cross-correlation between daughter of cell A (dA), and cell B, which have the niece-aunt relation (see Figure 15.b). In both cases, the predicted values become closer to the observed values after incorporating additional correlation in our model. Both plots contain N = 3942 pairs of cells. See also Figure 24.

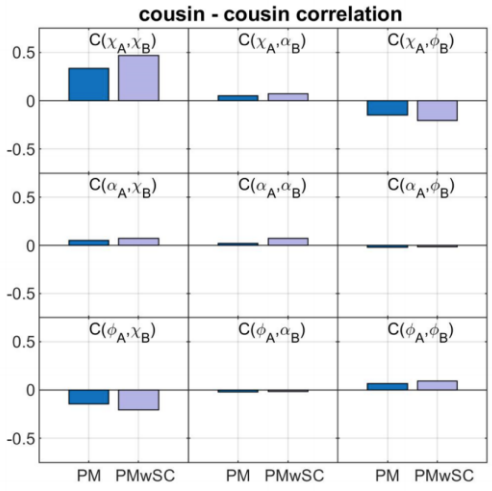


Figure 24 Prediction of the cross-correlation between two cousin cells. Our model allows for prediction of the cross-correlation between cousin cells (two cells whose mothers are sisters). Experimental measurement of these cross-correlations, however, is beyond the scope of this study and is subject to an ongoing investigation.

The additional correlation between sister cells can be readily incorporated into our model. The cross-correlation between N_A and N_B can be estimated from the data, as represented by a matrix shown in Figure 25. We find that the off-diagonal entries of the cross-correlation matrix are small. On the other hand, the first diagonal entry is -1, which means that N_A^{χ} and N_B^{χ} are fully anti-correlated as expected, since this reflects the fact that an increase in the birth size of a cell comes at the expense of its sister's birth size. This is due to the fact that the sum of the initial sizes of the two sister cells is equal to the final size of mother cell. Note that this is not contradictory to the results presented in Figure 23, which shows a positive correlation between χ_A and χ_B. This positive correlation reflects the fact that a larger mother will give rise to two larger daughters.

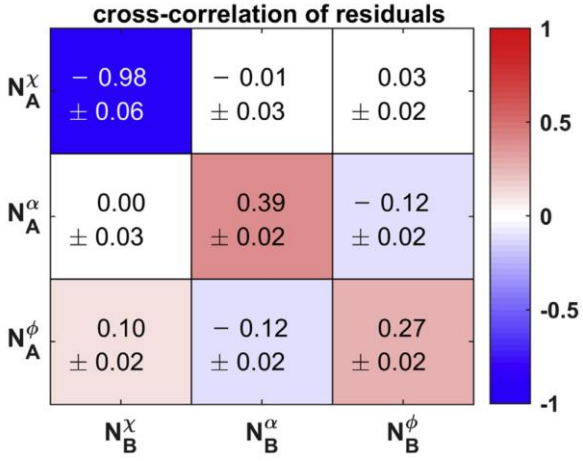


Figure 25 Additional cross-correlation between sister cells. Cross-correlation between the residuals N_A and N_B in Equation 3-5. The highest correlation is observed between N^{χ}_A and N^{χ}_B, which is associated with the birth length of the sisters (Equation 3-2). This value close to -1 is expected, since sister cells receive complementary amounts of birth size from their common mother. Other unexpected significant correlations are observed on the diagonal entries, $\langle N^{\alpha}_A N^{\alpha}_B \rangle$ and $\langle N^{\alpha}_A N^{\alpha}_B \rangle$ (see Equation 3-3 and Equation 3-4), which reveal new sources of correlation between sister cells. These correlations were calculated using data collected from N= 3942 pairs of sister cells.

More importantly, the other two diagonal entries, which represent the correlations between N_A^{α} and N_B^{α} and between N_A^{ϕ} and N_B^{ϕ}, are both positive. These two terms will contribute as new sources of correlation between the sister cells. Taking such additional correlations into account, we calculated new values for the cross-correlation between sister cells (Figure 23), PMwSC: predicted by the model with sister-cells correlations. These values show an improved agreement with the observed values.

3.9 Sister-cells correlation predicts correlation between niece and cousin cells

To test our generalized model with additional sister-cells correlation incorporated, we used the model to calculate correlations between cells that are genealogically more distant. For example, the model predicts a specific amount of correlation between a "niece" and an "aunt" cell. This correlation can also be calculated from the data, by comparing a cell (for example A[n]) with its "aunt" (B[n-1]) (Figure 15.b). Figure 23B shows the results of the predicted and the observed niece-aunt correlations. Note that existing cell growth models without incorporating the additional sister-cells correlation found in the previous section would still predict a small amount of correlation between niece and aunt cells (Figure 23B, PM), because these cells share a common ancestor. Incorporating the additional sister-cells correlation in the model improves the agreement between predicted and observed niece-aunt correlations (Figure 23B, PMwSC), supporting the existence of such additional sister-cells correlation that is missing from single-cell growth models.

Our model also allows us to make predictions about the correlation between two "cousin" cells, i.e., two cells whose mother cells are sisters. We calculated such cousin-cells correlation using our model with and without the additional sister-cells correlation (Figure 24). To

experimentally check the predicted cousin-cells correlation requires tracking the descendants of both sister cells. An experiment allowing such measurements is being devised by our group and will be reported in a future report. Notably, correlation between cousin cells has been reported before in mammalian cells, where cell-cycle time was found to be strongly correlated between cousin cells but not between mother and daughter[84]. In the newly devised experiment, we will measure such correlations in bacteria and further test our model of cell growth control.

4.0 Immediate bacterial growth correlation with cell content

Microorganisms live in a continuously changing environment, with fluctuations occurring both in time and space. Biological and biochemical processes that take place inside the cell and control biomass production and cellular growth are subject to molecular noise[85–87], and are strongly influenced by the environment[88]. Bacterial growth is determined by the rate of such biochemical reactions, their efficiency, and their collective organization. It plays an important role in maintaining cellular characteristics such as cell size, and it contributes to cellular fitness. Observations have indicated that bacterial cells are able to self-replicate accurately and within a well-defined timeframe. As explained in Chapter 1.0, cell growth and protein production increase exponentially during the replication timeframe. This type of growth, if subject to random noise, will lead to divergence in cell-size and protein content[12,89]. Therefore, to guarantee the reliability of the replication process, cells need to have mechanisms of "checks and balances" that will ensure tight regulation of cellular activities. Over the last few years, an extensive effort has been devoted to understanding the regulation mechanisms that control the cycle of microorganisms self-replication in various bacterial and yeast species[10,12,64,89–92]. In bacteria, as well as budding yeast, it has been shown that the cell size at the end of the cell cycle right before a division event is linearly proportional to the cell size at the start of the cell cycle, with the slope being ~1[10,93]. This suggests a model, in which cells add on average a constant size between two consecutive division events[10,12,94], known as the "adder" model (see Chapter 1.0). In a more recent work, it has been argued that the constant size is in fact added between two consecutive DNA replication-initiation events[93]. In yeast, the concentration of a cell-cycle regulator protein was shown to determine the division event and control cell size[90]. Such regulators, however, do not exist in bacteria, and to

date, the mechanism used to determine cell-size and growth rate, and regulate cell division in bacteria remains unknown.

In this chapter, we aim to connect the rate of growth to the cell composition. Many studies have made an attempt to identify the molecular mechanism of bacterial growth[95,96,97], but a fundamental understanding of what cell composition will lead to a maximum efficiency in growth, and how the different cellular components influence its rate are still unknown. To shed some light on this problem, we will investigate Equation 3-3 further, which reveals that the growth rate of a cell depends on the fraction of size it received from its mother. This equation suggests a mechanism of how two bacterial sister cells are regulating their size during the cycle right after their birth (see Sections 4.2 and 4.3) and could be a solution to develop a mechanistic understanding of cell size homeostasis. In Chapter 3.0[64], we examined the dynamics of cell size in long lineages of *E. coli* bacteria measured at the single-cell level in microfluidic traps. Our data were consistent with previous studies and agreed with the adder model at the population level (Figure 35a). However, comparison between sister cells in the same experiment, reveals that the smaller sister adds a larger size during the first generation after division (Figure 35b). These two graphs contradict each other at first glance, however, when examined in detail, uncover some new mechanisms explained in the following sections.

4.1 Materials and methods

4.1.1 Bacterial strains and plasmids

All experiments were performed using E. coli K-12 derivative MG1655. The following plasmids were used in this study. 1) PZA3R-mcherry, which is a medium copy number plasmid expressing a red fluorescent protein under the control of the λ-phage Pr promoter, 2) pRJ2001-GFP-Fis (a kind gift from J.F. Marko), where the GFP-Fis gene is under the control of the lac promoter[62], 3) pDB192-sulA+GFP-Fis, which was constructed by amplifying the GFP-Fis gene along with its promoter and terminator from the plasmid pRJ3306 (a kind gift from R. C. Johnson) by PCR using the primer (CGAGCTCCTCACTCATTAGGCACCC) that added the restriction site SacI to the 5' end of the DNA fragment, and the reverse primer (CCAAGCTTCGATCTTCTTTCCAGGCTTC) which added HindIII to the 3' end of it. This was then inserted between the SacI and HindIII sites of the plasmid pDB192-sulA (a kind gift from A. Amir[63]) to result in the Isopropyl β-D-1-thiogalactopyranoside (IPTG) inducible pDB192-sulA+GFP-Fis plasmid. This final plasmid induces both sulA and GFP-Fis under the lac promoter. A map of this plasmid is represented in Figure 26.

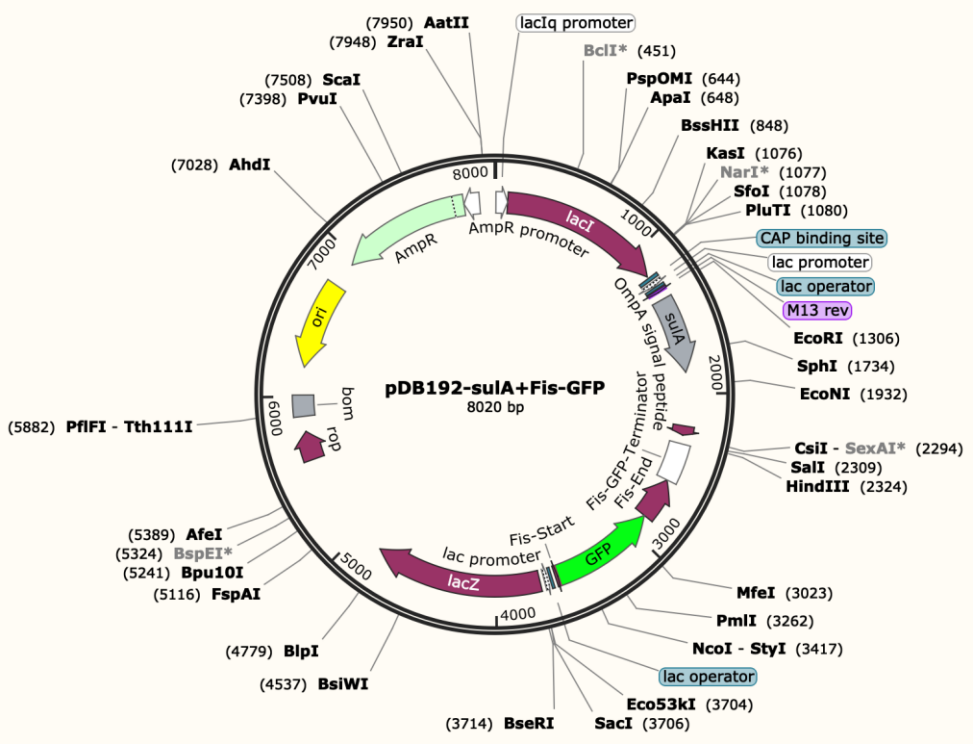


Figure 26 A map of the plasmid pDB192-sulA+GFP-Fis. This plasmid was constructed using the GFP-Gene from the plasmid PRJ3303 and inserting it into the plasmid pDB192-sulA.

4.1.2 Experimental setup and data acquisition

In experiments where cell filamentation was induced by exposure to one of two types of antibiotics, Nalidixic acid or Cephalexin, MG1655 *E. coli* cells containing two plasmids, PZA3R-mcherry, and PRJ2001-GFP-Fis, were used. The concentration of both nalidixic acid and cephalexin used was $5\mu g/ml$, and cells were exposed to the relevant antibiotic for 1.5 hours. All media getting pumped into the device contained 1mM IPTG for GFP-Fis induction. In experiments where filamentation was induced by overexpression of sulA, the same strain of cells was used but with the plasmids, PZA3R-mcherry, and pDB192-sulA+GFP-Fis. In these experiments, cells were exposed to 1mM IPTG for 2.5 hours to induce expression of sulA.

To acquire single-cell growth movies, cells were cultured overnight in LB medium at 32℃, and then diluted in the same medium the following day. When the cells reached Optical Density (OD) between 0.1 and 0.2, they were concentrated and loaded into a mother machine which was fabricated as explained in chapter 2.0. After sufficient number of cells were trapped, the excess cells were washed away with LB medium and the trapped cells were allowed to grow in the channels for one day. The next day, LB medium containing the desired antibiotic and/or inducer was pumped into the device for the desired time using a PTFE tube (Cole-Parmer, IL) with inner diameter of 0.022 inch and length of 0.7 meters at a rate of 1 ml/hr. With these conditions, it took the medium ~14.75 minutes to reach the device, so in all figures and graphs presented in this chapter, the time of adding or removing a chemical has been moved forward by 15 minutes. The medium was changed back to LB after this, and the cells were allowed to grow back to normal exponential growth for one day. Images were taken from the traps in DIC and fluorescence modes every 3 minutes during the entire experiment using a Hamamatsu ORCA-flash4.0 camera, mounted on a Nikon Eclipse T*i*2 inverted microscope with a 100X objective. An Okolab microscope enclosure stabilized the temperature at 32℃. Cell length and fluorescence intensity of cells were measured using Oufti[64] and custom-made MATLAB codes.

4.1.3 DNA concentration estimation

To find the true concentration of bacterial DNA using our experimental measurements, we consider the following chemical reaction: $[free\ GFP\ (fGFP)] + [free\ sites] \rightarrow [bound\ GFP\ (bGFP)]$, where [.] is the notation for concentration and in our case is fluorescence intensity. This reaction indicates that free GFP ($fGFP$) in the cytoplasm (GFP-Fis expressed by plasmid), binds to free binding sites on the chromosome, and produces Fis-DNA compounds,

namely bound GFP ($bGFP$). $[fGFP]$ and $[bGFP]$ are directly measurable from our experimental data, and $[free\ sites]$ is the value we are seeking to find. This reaction allows us to measure the true value of DNA concentration in cells that express GFP-Fis as follows. Considering the above chemical reaction goes to equilibrium in time scales much smaller than our measurement time[65], we can use the law of mass action: $k_a[fGFP][free\ sites] = k_d[bGFP]$, where k_a and k_d are the association and dissociation constants, respectively.

We find the concentration of free GFP, $[fGFP]$, by measuring total GFP intensity in multiple small areas which are located outside the chromosome but inside the cytoplasm and divide it by the sum of areas we have selected (Figure 27). Figure 28 illustrates sample measurements for $[fGFP]$ in single cells during their filamented cycle. On the other hand, concentration of bound GFP, $[bGFP]$, is equal to the total intensity of chromosome divided by cell area. We use two intensity measurements from our experiments to find this. We subtract the total GFP of the cell by the total free GFP of the cell to get total bound GFP. In other words, $[bGFP] = \frac{total\ GFP - [fGFP] * cell\ area}{cell\ area}$. Figure 29 illustrates sample measurements for $[bGFP]$ in the same single cells as in Figure 28. The measured concentration of bound GFP and free GFP in cephalexin, nalidixic acid, and sulA are presented in Figure 30, Figure 31, and Figure 32, respectively. As it can be seen from these figures, $[bGFP]$ does not change in any of these cases, which is a demonstration that all free sites on the DNA are bound to Fis molecules at all times. However, $[fGFP]$ changes only in the case of nalidixic acid. This could be due to the fact that in this case, the DNA is not being replicated and so there are more free molecules of Fis available in the cell.

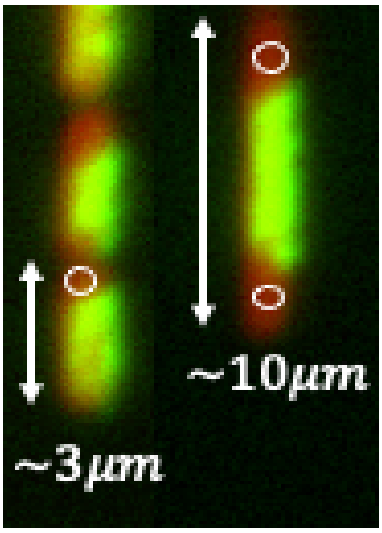


Figure 27 A sample cell in two different time points when exposed to nalidixic acid. In the first frame, the cell is in the beginning of its cycle and has a length of $\sim 3\mu m$, and in the second frame the cell has elongated to a length of $\sim 10\mu m$. The color red is mcherry expressed from the plasmid PZA3R-mcherry, which was used as a sample protein expression measure, and the color green is GFP expressed from the plasmid PRJ2001-GFP-Fis. The intensity of green is much higher when Fis molecules are bound to the chromosome, but some can also be detected in the cytoplasm area when measuring green intensity in white circles. By averaging this value we obtain $[fGFP]$. $[bGFP]$ is subsequently found using the method described in the text.

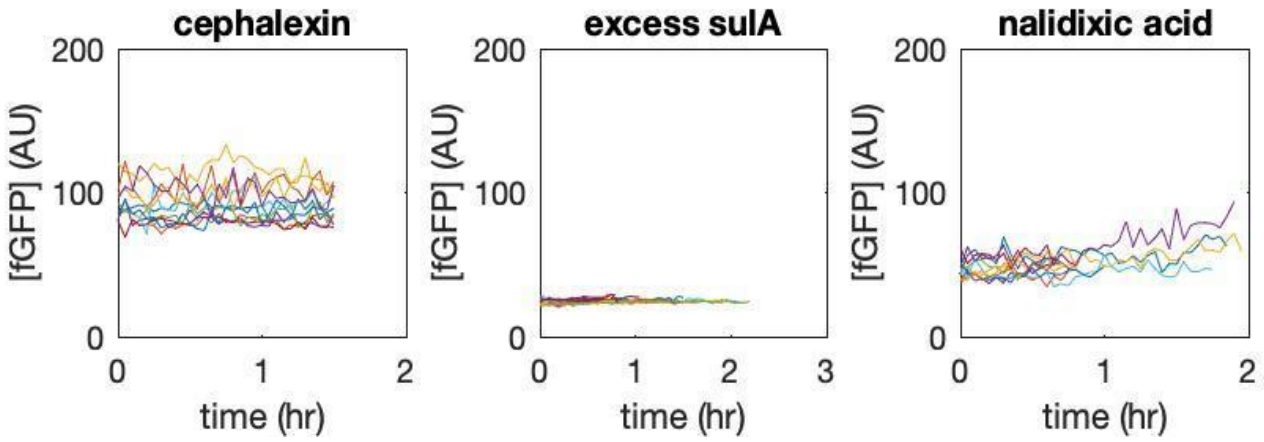


Figure 28 Concentration of fGFP in sample single cells when exposed to cephalexin, excess sulA and nalidixic acid. Each line represents one single cell.

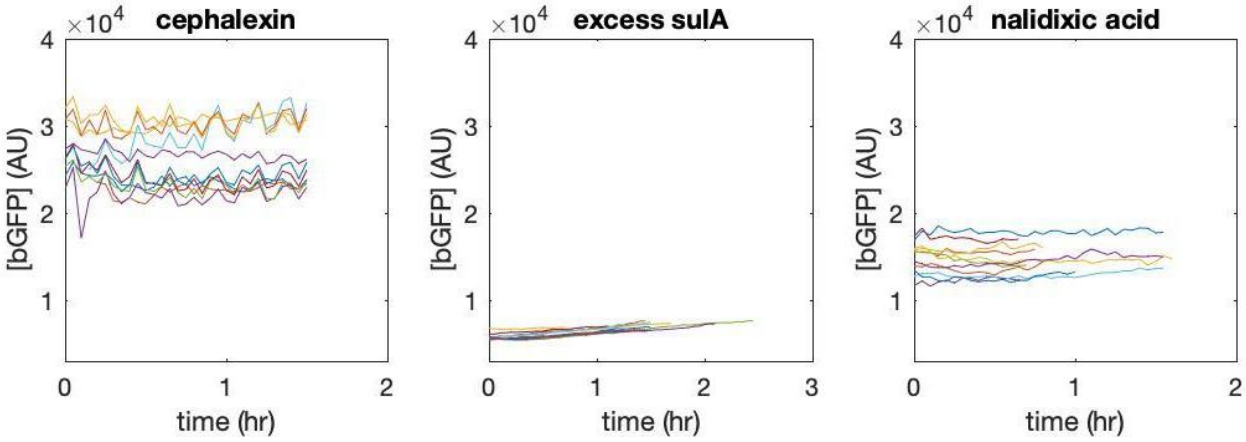


Figure 29 Concentration of bGFP in the same sample single cells as presented in Figure 28 when exposed to cephalexin, excess sulA and nalidixic acid. Each line represents one single cell.

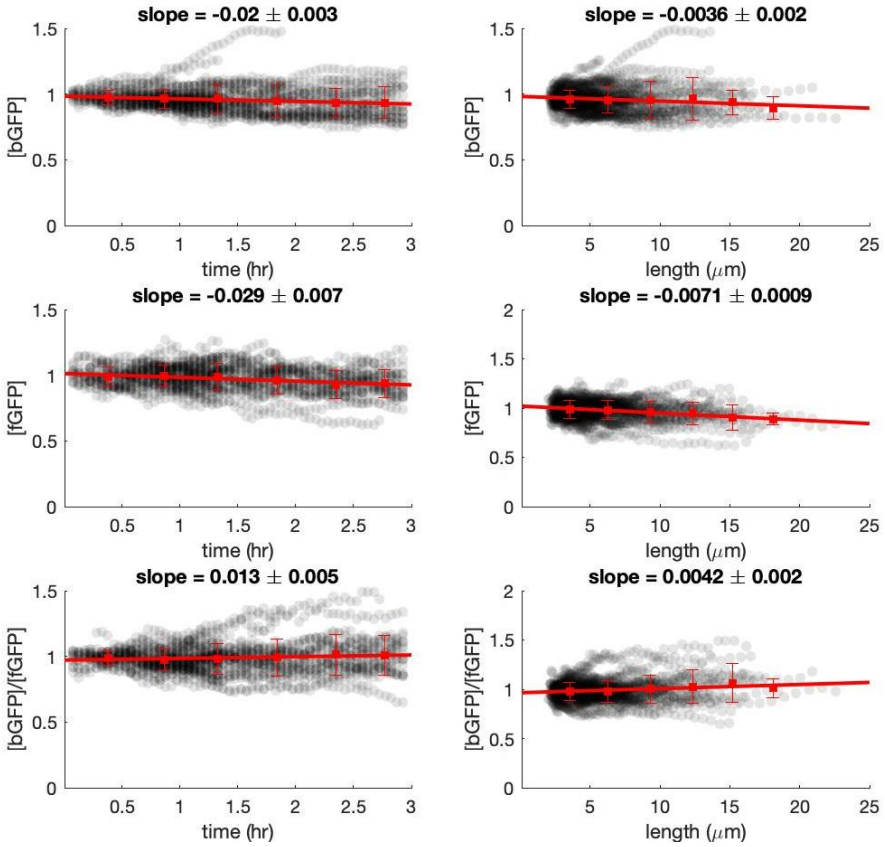


Figure 30 Concentration of bound and free GFP in the experiments where cells were exposed to cephalexin. Black dots are measurements from individual cells, red squares are binned data, and red line is a fit to these points.

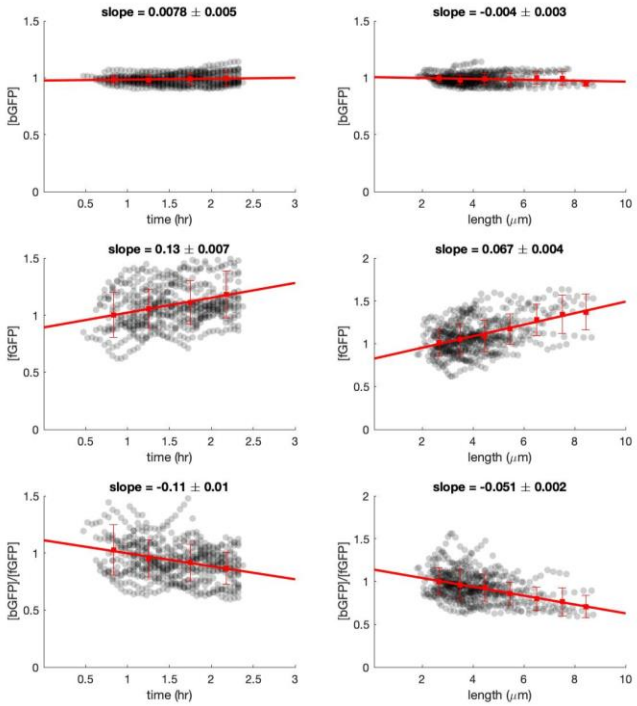


Figure 31 Concentration of bound and free GFP in the experiments with nalidixic acid. Black dots are measurements from individual cells, red squares are binned data, and red line is a fit to these points.

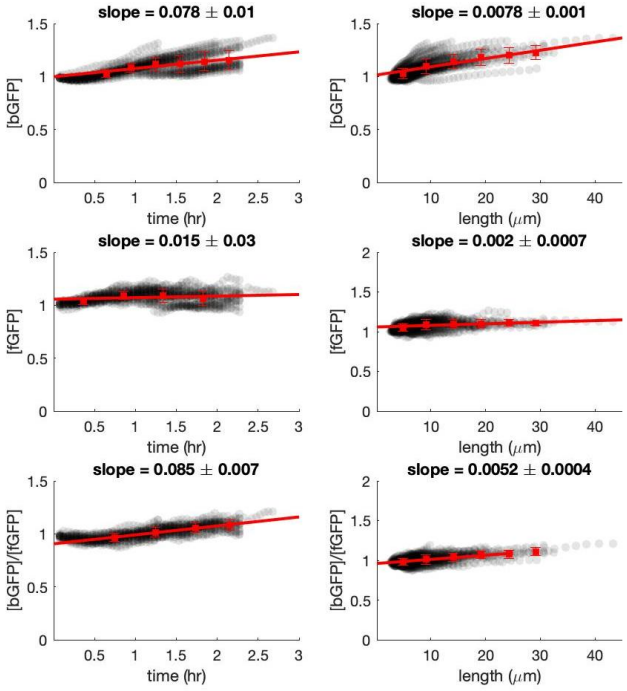


Figure 32 Concentration of bound and free GFP in the cells with excess sulA. Black dots are measurements from individual cells, red squares are binned data, and red line is a fit to these points.

In conclusion, we can find the concentration of free binding sites by using $[free\ sites] = c_0 \frac{\frac{tot\ GFP - [fGFP] * cell\ area}{cell\ area}}{[fGFP]}$, where c_0 is a constant equal to $\frac{k_d}{k_a}$. In this equation, the concentration of free sites is related to the concentration of DNA, $[DNA]$. We can write this as: $[DNA] = c * [free\ sites]$, assuming that Fis binds to DNA in a sequence-independent manner, and in the case of additional Fis molecules, they bind to previous Fis-DNA filaments cooperatively and create higher-order complexes[66]. The value of the constant c is not in our interest, so we use the calculated value of $[free\ sites]$ as a direct measure for DNA concentration.

4.1.4 Fluorescence calibration experiments

Binding of Fis to the chromosome of E. coli could change due to the antibiotics that were used in our experiments, which make the shape of the chromosome to change during growth. however, DAPI is a dye whose binding to DNA does not change with the change of DNA shape, so the relation between the fluorescence of DAPI and DNA concentration is linear[98]. To ensure that Fis also binds to DNA in a similar form and that the concentrations of DNA that we calculated are reliable, we used our strain, MG1655 with two plasmids PRJ2001-GFP-Fis and PZA3R-mcherry and started an experiment similar to the ones described in Section 4.1.2, with the difference that the experiment was set up under a Zeiss Axio Observer microscope with a 100x objective. $5\mu g/ml$ nalidixic acid was added to the medium and left to flow through the device for 3.5 hours. One set of images were taken from a total of 190 channels which contained 837 single cells using four modes, phase contrast, green and red fluorescence, and UV. Then the device was washed using 1X PBS (8 gr NaCl, 0.2 gr KCl, 1.44 gr Na_2HPO_4, 0.24 gr KH_2PO_4, 1 L DI water, pH=7.4), and the bacteria were fixed by being exposed to 10% formalin (100 mL 36.5-38.0%

formaldehyde, 4 gr NaH_2PO_4, 6.5 gr Na_2HPO_4, 900 mL DI water) for 20 minutes[99]. At this point, 1X PBS containing 3 μM DAPI was pumped into the device for 30 minutes. The device was washed again with 1X PBS and left to stabilize for 20 minutes. The second set of images was taken at this point, again using all four channels. The temperature was fixed at 32℃ using a made-in-house incubator at all times. The total and free green fluorescence and UV were measured from the images similar to explained in the previous section, and the concentration of each channel was calculated similarly ($[DAPI] = \frac{\frac{totUV-([fUV]\times area)}{area}}{[fUV]}$). The calculated values for concentration of UV and GFP were plotted against each other (Figure 33) and a linear relationship between them was observed. This ensures that the concentrations calculated in Section 4.1.3 are reliable.

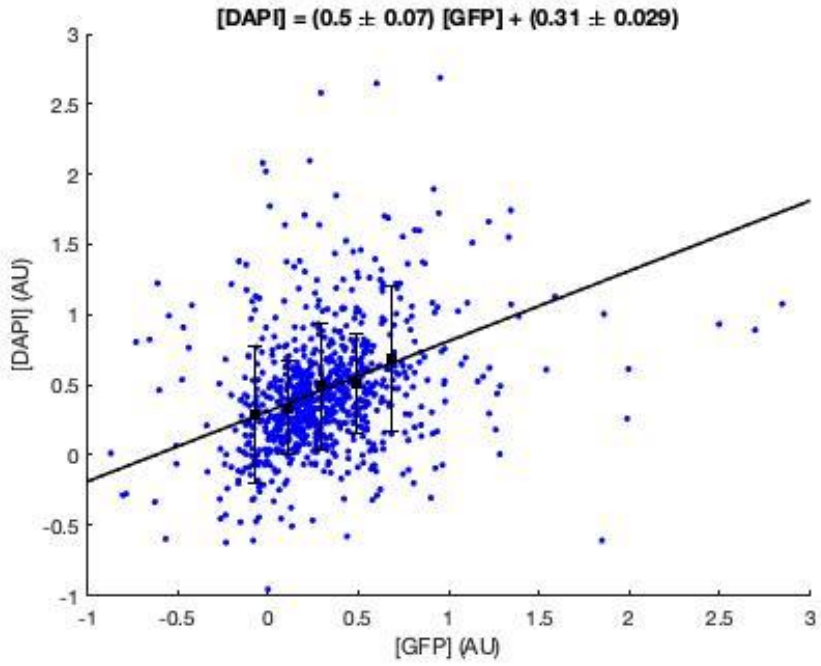


Figure 33 The emission of UV light from the dye DAPI as well as the fluorescence emission from a green protein connected to Fis were measured and plotted against each other to make sure that binding of Fis molecules does not change due to the addition of the antibiotic nalidixic acid. A linear relationship between them ensures the calculations performed in Section 4.1.3 are reliable.

4.2 Division of *E. coli* is symmetric

Let us start by stating that each cell-cycle, can be well described with the following mapping model:

$$x_n(t) = x_n(0)e^{\alpha_n t}, \quad 0 \le t \le T_n$$

$$x_{n+1}(0) = f_n x_n(T_n) \quad \rightarrow \; x_{n+1}(0) = f_n x_n(0)e^{\alpha_n T_n}$$

where x_n is cell size at the n-th cycle, T_n the duration of the n-th cycle, α_n the exponential growth rate, and f_n the division ratio. This equation links the cell size at the start of the cell cycle along generations. In this discrete mapping, cell size undergoes a total increase by a factor of e^{φ_n} and a decrease by a factor of f_n every cycle. φ_n is the total accumulation exponent during the n-th cycle which is connected to the growth rate and cycle time as $\varphi_n = \alpha_n T_n$, where f_{n+1} is the size fraction that a daughter cell receives from its mother. When measured from experimental data, the distribution of this parameter can be well-fitted to a Gaussian plot and will have a mean of 0.5 as can be seen in Figure 34. This distribution suggests that cell division is symmetric with noise and has been therefore assumed not to affect growth rate of the cell. In chapter 3.0, Equation 3-2, we found that the size fraction is random and independent of other variables, and we explained how this independence can be understood from the complementarity of the daughter cells size-fractions to each other. In short, since Equation 3-2 should be applicable to both daughter cells, and the sum of their size-fractions should be 1, neither of the fractions can depend on the variables of the mother cell. However, this does not mean that the variables of daughter cells do not have a dependency on the fraction. In fact, our experimental analyses reveal that there is a dependency of some daughter variables on the fraction that it receives from its mother.

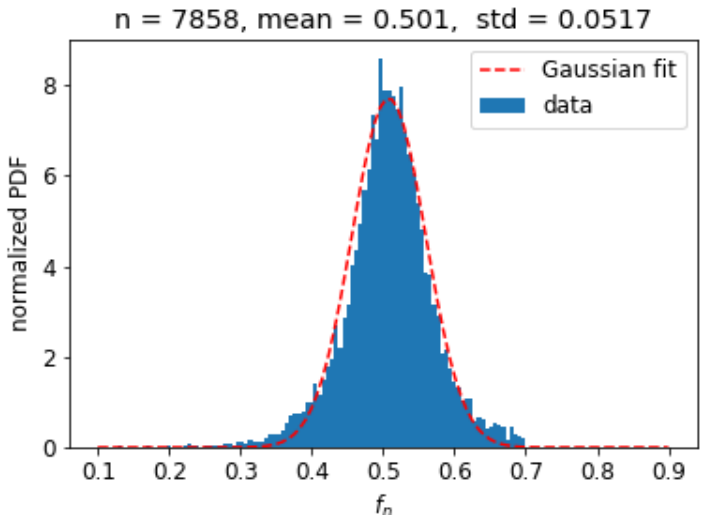


Figure 34 Distribution of fraction of size that daughter cell receives from mother cell. The Gaussian fit ensures that cell division is in fact symmetric and an independent variable, however it does not imply that other variables are independent of fraction.

The adder model, explained in Chapter 1.0, suggests that the amount of size that a cell adds in each cycle is constant and independent of the cell size at the start of the cell cycle. This is a phenomenon which to date is found to be true for all experimental data acquired from *E. coli* cells, however, it does not explain the mechanism of cell growth. Figure 35a presents the amount of cell growth (Δ) plotted against birth size ($x(0)$) and it is evident that the adder phenomenon is consistent with our single-cell data as well. This clearly stands in contradiction to what we found in Equation 3-3, which reveals that the growth rate of a cell depends on the size-fraction that it receives from its mother at birth. The cells that were used to plot this graph, if from the same experiment, all have the same ancestor (they were grown from a single colony from an agar plate), but their relationship is very distant, and this causes the phenotypes to have a huge variability. In Figure 35a this factor has been ignored, along with other factors, such as possible variations among cells measured in different experiments, and cell history which might have long-term effect on cellular dynamics. To eliminate these elements and reduce variation among cells considered, we

examined the difference between two sister cells immediately after separation from a single mother. smaller variation is expected between these two cells since the mother divides almost symmetrically. We observe a dependency of the difference in the added sizes of the two sister cells on the size-fraction they receive from their mother. Figure 35b demonstrates this dependency where a and b denote the two sister cells and are assigned randomly. The large negative slope of the linear fit to the binned data, points to a new growth control mechanism that has not been considered previously. This inverse dependency of the added cell size on the size-fraction received, implies that the smaller sister is adding a larger volume during the cycle immediately following its birth compared to its larger sister.

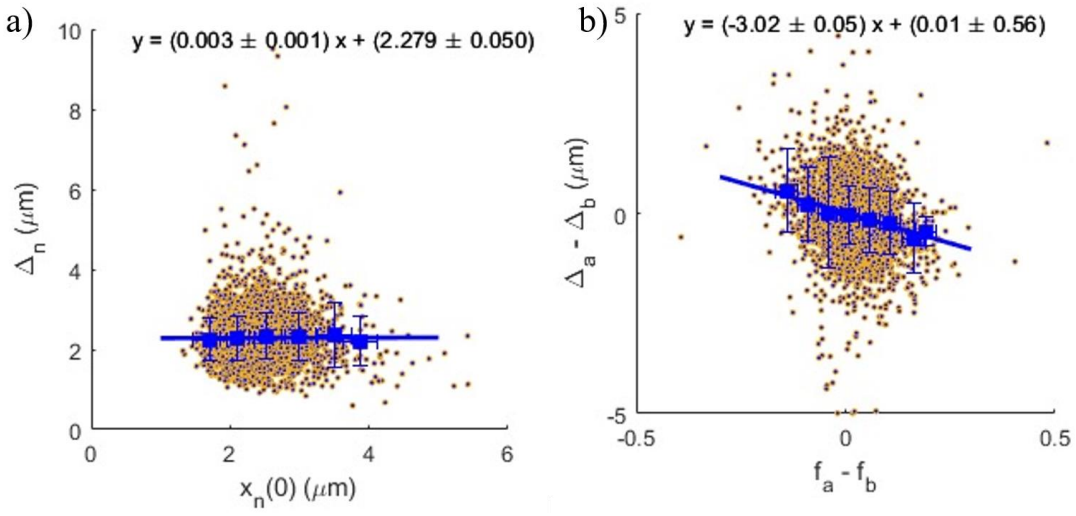


Figure 35 A hidden growth mechanism is uncovered by removing the huge variability in a population of cells. a) The amount of length added (Δ) during a cycle (n) versus the birth length ($x(0)$) of the same cycle. Blue squares are binned data and blue line is a fit to the bins. The horizontal line with a small slope (0.003) indicates no dependency between these two variables. b) The amount of length added in one daughter cell (Δ_a) subtracted by the amount added by its immediate sister (Δ_b) versus the difference in fractions that they receive from the mother ($f_a - f_b$). Cells a and b are randomly apponted to the sister cells. The high negative slope (-3.02) indicates a dependency between these two parameters.

4.3 Smaller sister has a higher growth rate

The parameter Δ_n in Figure 35 is in fact proportional to the total length accumulation, $\phi_n(T_n)$, where n is the cycle number, and T is the cycle time. As denoted before, length accumulation itself has a dependency on two different measurable parameters: $\phi_n(T_n) = \alpha_n T_n$, where α is the growth rate of the cycle. To determine which of these parameters contributes to the observed difference in added volume between the two sister cells, we examined the distributions of the two parameters and their differences between the two cells. We denote the smaller sister as "a" and the larger sisters as "b". Our results, presented in Figure 36, show a zero mean for the difference between cycle times of the two sisters, but a slightly positive mean for the difference between their growth rates. It can be concluded that sister "a" (the smaller sister) has a higher growth rate compared to its larger sister and both sisters grow for a similar time period.

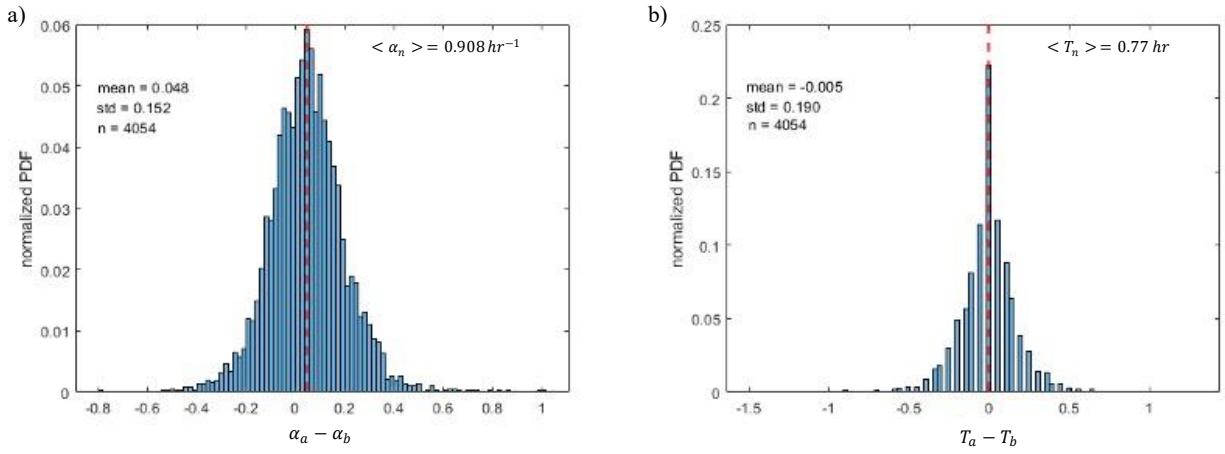


Figure 36 Distribution of difference in two parameters in sister cells. "a" is always denoted to the smaller sister and "b" to the larger sister. a) The difference between growth rate of sisters exhibits a slightly positive mean indicating that the smaller sister has a higher growth rate. b) The difference between cycle time of sisters has a mean value very close to zero indicating cycle time is the same, independent of the fraction.

To further verify this result, in Figure 37 we inspect these differences as a function of the difference in the fractions both cells receive from their mother. Figure 37a shows that the cell cycle time is independent of the size fraction the cell receives from its mother. On the other hand, the large negative slope of the linear fit to the data in Figure 37b suggests an inverse dependency of the cell's growth rate on the size fraction it receives from its mother. We thus conclude that the sister that receives a smaller size fraction from the mother will grow faster than its larger sister. This is consistent with the results presented earlier in Chapter 3.0.

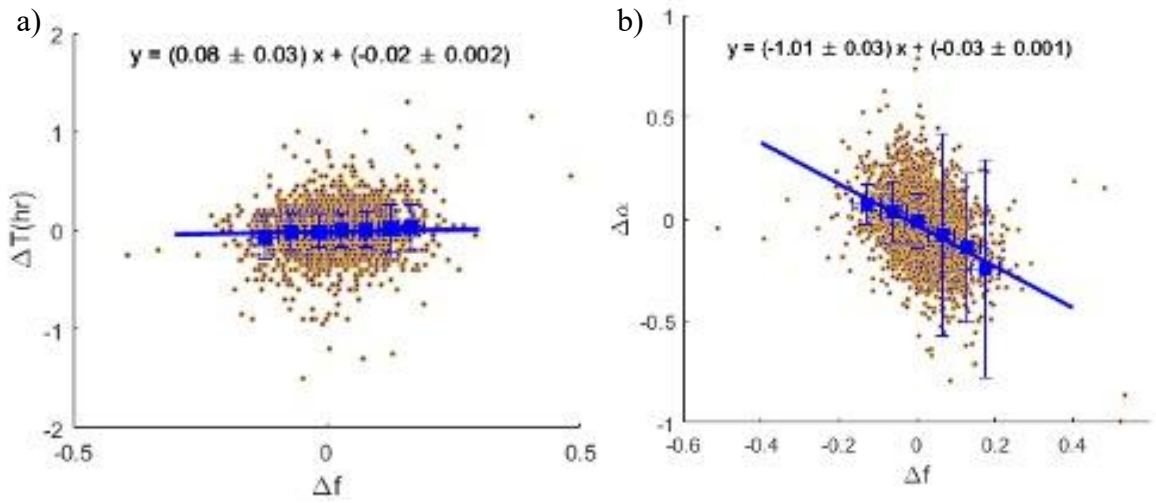


Figure 37 Dependency between sister parameters and size fraction, where "a" is the smaller sister and "b" is the larger sister. a) $\Delta T = T_a - T_b$ and $\Delta f = f_a - f_b$. Blue squares are binned data and the blue line is the fitted line to these points. The slope close to zero (0.08) indicates no dependency between cycle time and fraction. b) $\Delta\alpha = \alpha_a - \alpha_b$ and $\Delta f = f_a - f_b$. The high negative slope (-1.01) suggests a negative dependency between growth rate and size fraction, interpreted as the smaller sister having a higher growth rate.

4.4 Growth rate varies in one cycle

Previous studies always assumed that the growth rate of a cell is constant during one cycle (but varies from cycle to cycle). Therefore, the growth rate was always determined using exponential regression over the size measurements of a complete cell cycle as displayed in Figure 38a. The results of Section 4.3, however, reveal that when the cellular content between two cells is almost identical then the growth rate depends on the cell size at the start of the cell cycle. This results in the smaller cell increasing its volume faster than the larger cell. If the growth rate dependence on the cell size is maintained throughout the cell cycle, then it would lead to smaller difference in the growth rate between the sisters as the cell cycle progress.

To test this hypothesis, we evaluated the growth rate during various periods of the cell cycle separately and independently of each other. Instead of fitting all of the complete cell cycle to one exponent, we applied the exponential regression to a moving window of five points at a time, which constitute ~half of a cell cycle. This provides a number of exponents for each cycle as depicted in Figure 38b-c. We applied this fitting method to both sister cells, and for many pairs of sisters. We then evaluated the difference in these exponents between sisters, whose average is depicted in Figure 39 . $\Delta\alpha$ in this graph is always the growth rate of the larger sister subtracted from that of the smaller sister. Similar results were also obtained in another study of our group using different methods[100]. Surprisingly, this graph reveals that at the beginning of a cycle, the difference between the growth rates of sisters is high, however, as time goes by and they get closer to the middle and the end of the cycle, this difference becomes smaller until it approaches zero. We interpret this into a growth mechanism that when a mother cell divides into two daughters, the smaller one grows faster in the beginning, which allows the smaller cell to compensate for the size

difference acquired during division, and as the size difference becomes smaller the growth rates become similar (Figure 40).

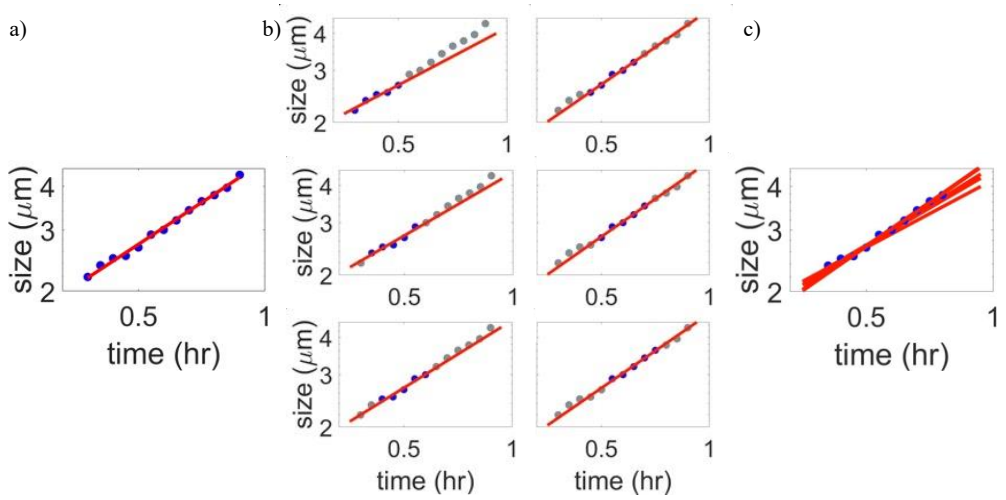


Figure 38 Exponential regression on cell size. Blue dots are measured values and the red line is an exponential fit. a) So far, we have been fitting all the points of one cycle to one exponent and appointed that as the growth rate of the cycle. b) Exponential fitting was applied on a window of five points, then the window was moved by one point until the last point of the cycle was reached. One exponent was found for each fit. c) All exponential fits for one cycle presented in a single graph.

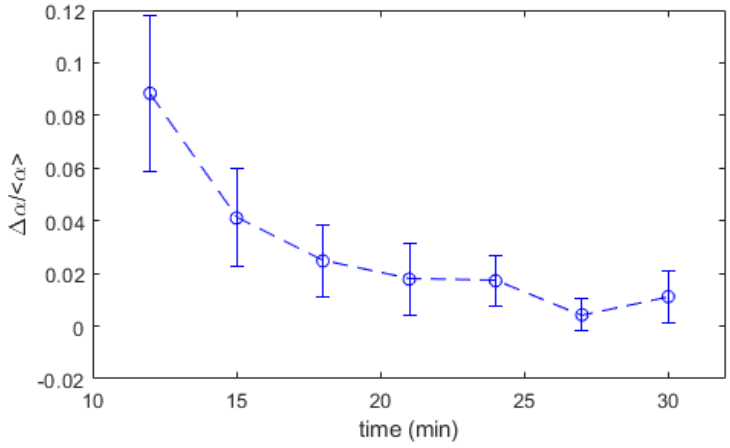


Figure 39 Difference in growth rates between sisters in one cycle. $\Delta\alpha = \alpha_a - \alpha_b$, and is the average of many pairs. a is always the sister with a smaller birth size and b always the larger one. This difference has been normalized by dividing by the average of the growth rate of all cells (both sisters). Error bars are standard deviation.

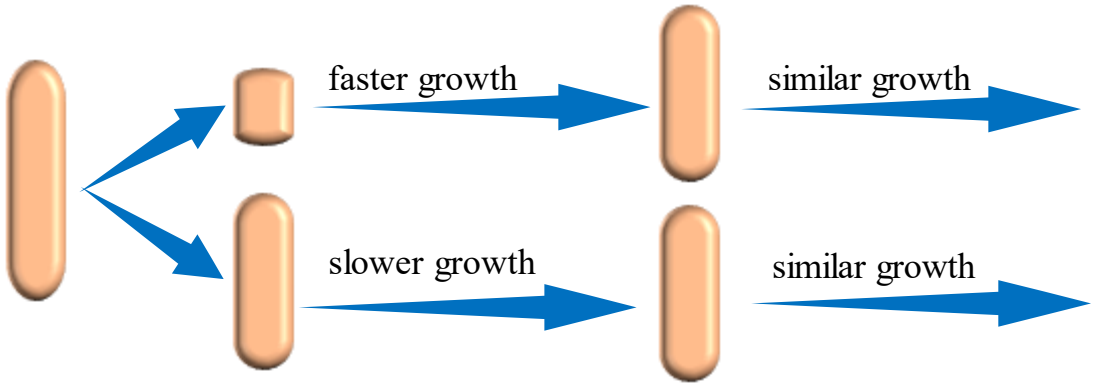


Figure 40 Our analyses are interpreted into a growth mechinsm that when a mother cell divides into two daughters with a slight difference in size, the smaller one grows faster than the larger one in the beginning of the cycle. After the smaller one has compensated for its size difference, their growth rate becomes similar.

4.5 DNA concentration is unequally distributed in sisters

Our analyses so far demonstrate that indeed the difference in the growth rate of two sister cells at the beginning of their cycle is proportional to their difference in size fraction as illustrated in Figure 40. This result offers an important insight into a possible source of the feedback responsible for size homeostasis. It suggests that whatever controls the growth rate to make up for the size inequality between two sister cells during division is inversely proportional to the size fraction the cell receives from its mother. Note that, in our case, it is not the *absolute* size of the daughter, but rather the *relative* size compared to the mother, that affects the growth rate. This is a very important distinction. It implies that in addition to the growth rate of the mother cell (i.e., the contribution of α_{n-1}), the distribution of resources between cells during division is also critical. This raises the question: what resource is distributed unequally between sister cells and can result in growth rate variation? To answer this question, we should remember that the rate of cell growth

depends on the rate of biochemical activity, which is determined by the concentration of resources and not their amount. Most resources in the cell, including enzymes, ribosomes, amino acids, and the various chemicals and proteins, are expected to be distributed between sister cells with equal concentration with some random noise that will not lead to a measured bias. Even membrane proteins are mostly distributed with equal concentration between the cells, and therefore cannot be responsible for the observed difference in growth rate. DNA is the only resource that is distributed between two cells in equal amounts (same number of copies) and not concentration. Thus, a cell with a smaller size fraction will have a higher concentration of DNA. This does not mean that the macromolecular crowding is higher in the smaller cell, but that the proportionality of the DNA to other molecules is higher. Such mechanism can provide a feedback source to offset size difference between cells, which is fundamental for size homeostasis.

To test this hypothesis, our former data are not sufficient, due to the small dynamical range that cells have while growing in desirable environmental conditions. The difference between the fraction that sisters receive, and their DNA concentration in these conditions is insignificant and not detectable. Therefore, we developed a new experimental procedure in which the dynamical range of cellular growth is broadened to allow better evaluation of the effect of DNA concentration on the growth rate. In these new experiments, cells were made to filament with and without DNA replication. In both cases, protein production and cell growth continued, and the only difference was the number of DNA replicates in the cell. This was facilitated by exposing cells to a very low concentration of antibiotics while growing in the microfluidic traps. Nalidixic acid is a quinolone antibacterial agent which results in DNA strand breaks. The outcome of this is either a decrease, or complete inhibition of DNA synthesis, causing elongation in a cell without division[101]. Thus, cells subjected to nalidixic acid would grow to lengths of ~20μm and continue to produce proteins

as displayed in Figure 41A. It can be seen that the DNA in the cell exposed to nalidixic acid is not segregating during elongation and only one nucleoid exists in the cell. While our measurements of the DNA concentration in cells exposed to nalidixic acid (Figure 49) do not conclusively determine whether the DNA synthesis is completely inhibited or is only slowed down, they do show a significant decrease in concentration. On the other hand, subjecting the cells to a small amount of cephalexin prevents the cell from dividing by interfering with the septum of cell division, while protein is continuously being expressed, DNA is being replicated and the cell is growing in size[102], as presented in Figure 41B. In this figure it is evident that several nucleoids exist in one elongated cell which is exposed to cephalexin. Previous studies have observed the effect of both of these antibiotics on populations of wild type *E. coli* cells[103,104] and some of its mutants[105], here we focus on their effect on the DNA replication in wild type bacteria, and how that influences the cell's growth rate at the single-cell level. Figure 42 illustrates sample traces of two cells where each one was exposed to one of the antibiotics. Here, a red fluorescent protein expressed from a medium copy number plasmid was used as a sample protein measurement, and a green fluorescent protein attached to the protein Fis (which binds to the chromosome in a non-specific manner) was used to measure the DNA concentration. For details on how DNA concentration was estimated in these experiments see Section 4.1.3. Except for the type of antibiotics introduced to induce filamentation, all other experimental conditions and procedures were kept identical between the two experiments (see Section 4.1 for more detail).

Our results show that while cells filament in both cases, their growth rate remains exponential similar to their growth before the addition and after the removal of antibiotics, only when exposed to cephalexin (with DNA replication) as seen in Figure 42b and Figure 43a. Figure 42a and Figure 43b, on the other hand, show that cells grow exponentially prior to exposure to

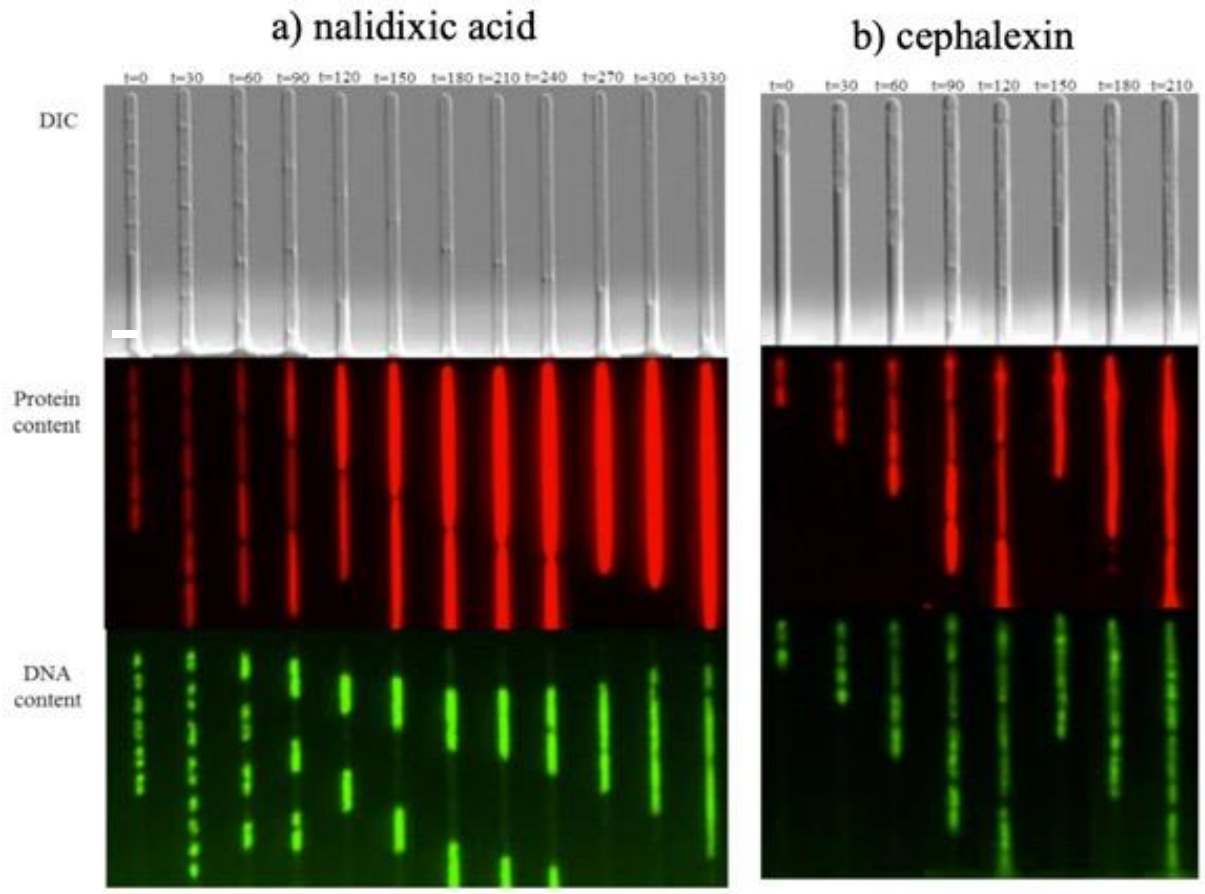


Figure 41 Cell filamentation with and without DNA replication. The color red is a red flourescenc protein (mcherry) expressed under the control of the λ-Pr promoter. The green images depict the protein GFP-Fis, which bind to DNA and are used to estimate the DNA concentration (see Section 4.1.3). (A) Nalidixic acid prevents cell division and induces filamentation by blocking the replication of DNA as can be seen in the green images, which show a single nucleoid region as the cell grows. (B) Cephalexin induces filamentation while allowing DNA to replicate as it is clear from the visible multiple nucleoid regions in green images. In both cases the cells continue to produce protein as can be seen in red images. White line on the top is 2 μm.

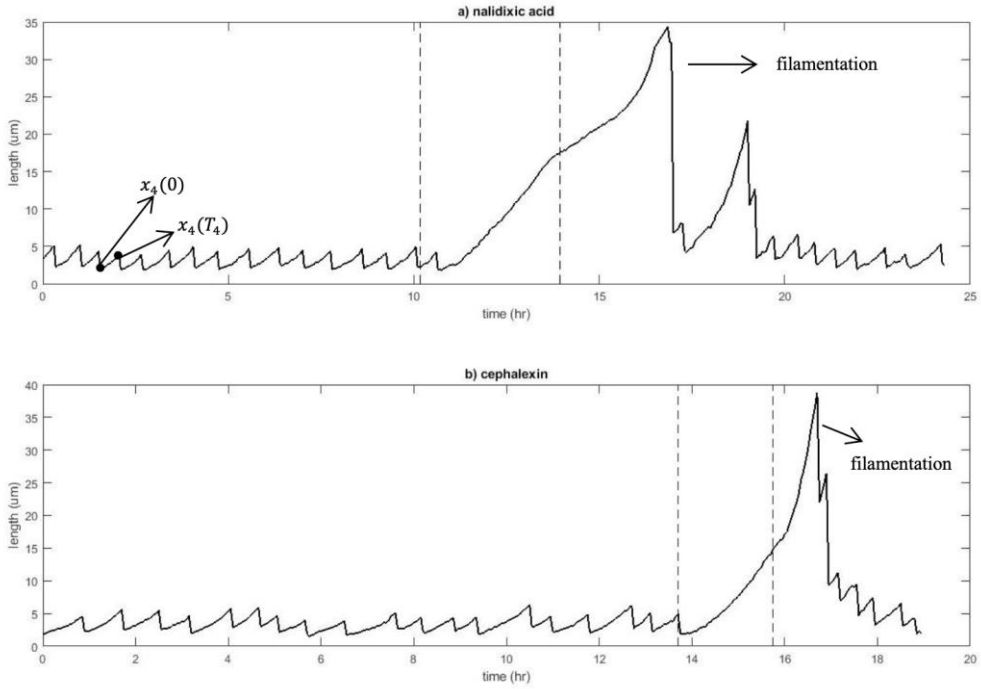


Figure 42 Sample trajectories of wild type *E. coli* cells growing in the absence and in the presence of a) nalidixic acid, and b) cephalexin. The cell was exposed to the antibiotic between two vertical dotted lines. After removal of the antibiotic, the cell goes back to normal exponential growth, ensuring both antibiotics are reversible.

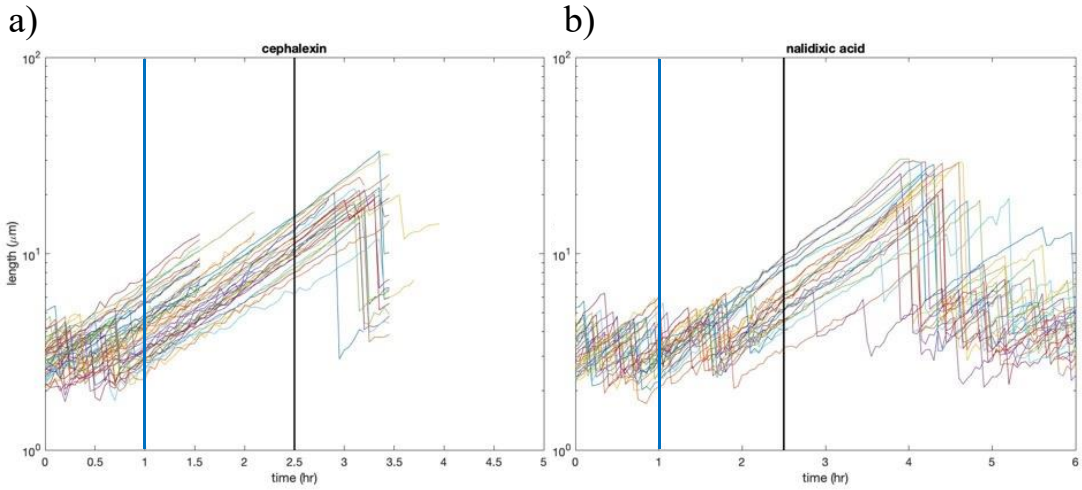


Figure 43 Growth of single cells when exposed to two different antibiotics. The antibiotic was added at time 1hr indicated with a blue vertical line, and removed at tim 2.5hrs indicated with a vertical black line. a) Growth is exponential during and after exposure of cephalexin because DNA is being replicated as in normal

growth. b) Growth deviates from exponential curve when exposed to nalidixic acid, but returns to exponential growth after removal of the antibiotic. It can be seen that the lines in (a) are more straight compared to lines in (b), which is the indication of exponential growth with cephalexin and non-exponential growth with NA. A more clear comparison is presented in Figure 44.

nalidixic acid, deviates from an exponential growth upon exposure to nalidixic acid (no DNA replication) and goes back to exponential growth after its removal. These figures also reveal that the cells divide once after addition of the antibiotic and before filamentation, which could be due to the fact that multiple chromosome copies always exist in a cell, so they are still able to go through one more division before the number of completed chromosome copies is too low to allow division. Furthermore, after removal of nalidixic acid, it takes around one hour for the cells to resume the exponential growth. This delay could be due to the fact that it takes some time to replicate the DNA before the cell is able to divide again. This delay is not observed with cephalexin since in that case growth is always exponential. The growth of a few sample cells when DNA is being replicated (cephalexin) and when DNA replication is inhibited (nalidixic acid) are compared in Figure 44. The slowdown witnessed in the blue graphs is consistent with a population growth model with carrying capacity[106]. In this model, cells in a population have to compete for resources, and therefore when the resources are limited, the growth rate will be determined by the amount of resources available per cell. As the population grows, this amount decreases, and as a result so does the population growth rate. Similarly, if that resource is the DNA, as the cell grows larger, the competition for the DNA increases and therefore the rate of protein production and cell growth decreases. Note that it has already been shown that the concentration of RNAP affects the growth rate of the cell[95]. However, the concentration of RNAP cannot account for the variation in growth rate among sister cells as it is divided equally between the two.

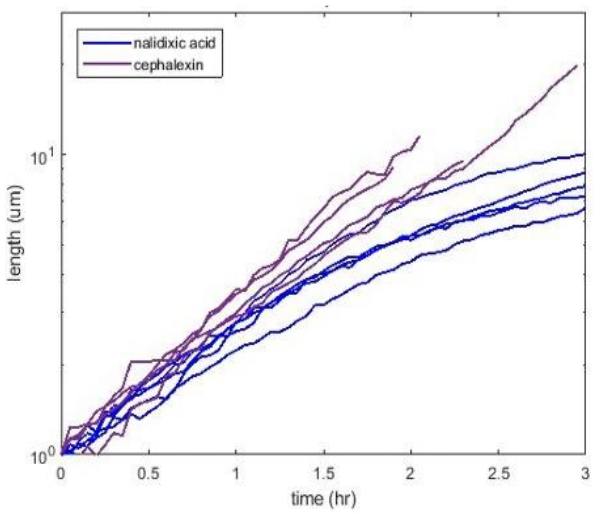


Figure 44 Comparison between growth of cells when DNA is being replicated (exposure to cephalexin – purple lines) and when DNA replication is inhibited (exposure to nalidixic acid – blue lines). It is apparent that when DNA is consistently being replicated growth remains exponential, but a slowdown is witnessed when DNA replication is inhibited.

4.6 Lower DNA concentration causes lower growth rate

The growth rate of single cells exposed to both antibiotics was calculated in order to verify and quantify the slowdown observed in Figure 43. As mentioned before, we did not fit the entire cycle to one curve, but to several curves. The growth rate (alpha) was calculated in the same manner explained in Section 4.4 and Figure 38, by fitting an exponential function to five points at a time. The results of such analysis are presented in Figure 45 and Figure 46, plotting once as a function of time and once as a function of length, for growth with and without DNA synthesis, respectively. These figures reveal that the growth rate of cells remains constant under cephalexin (with DNA replication) but slows down drastically under Nalidixic acid (no DNA replication). These values indicate no dependency of the growth rate of cells on time or length where DNA

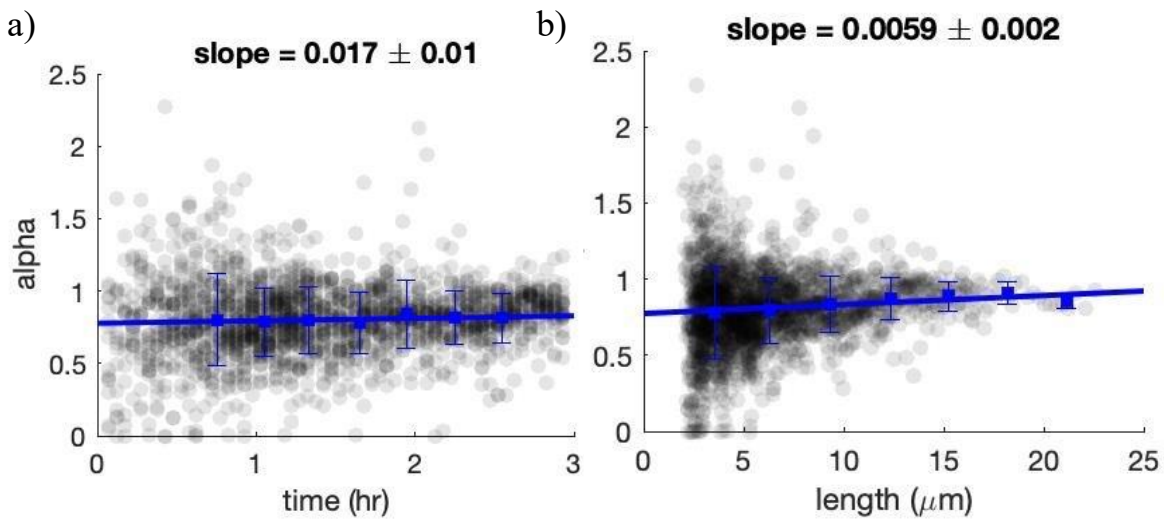


Figure 45 Growth rate (alpha) of single cells exposed to cephalexin as a function of a) time and b) length. Black dots are single points from 45 cells, blue squares are binned data, and blue line is a linear regression to the binned data. Error bars are standard deviation.

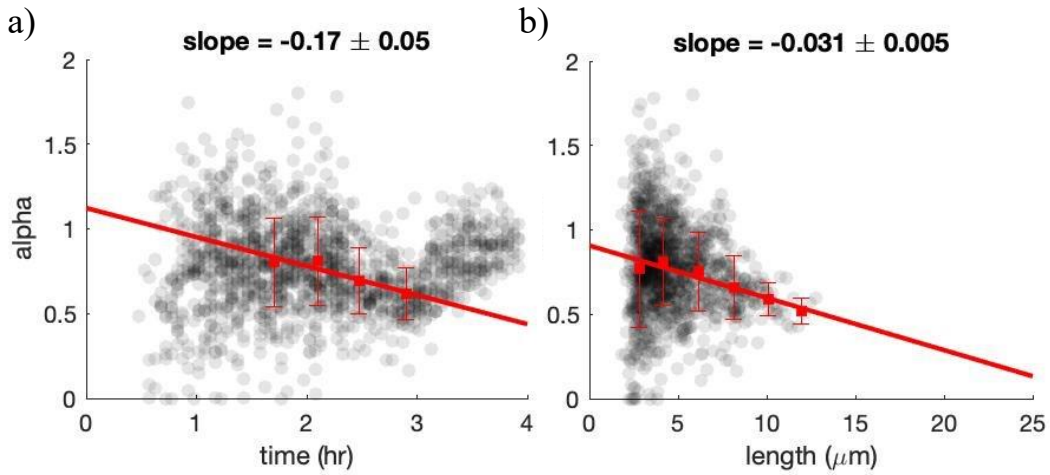


Figure 46 a) Growth rate (alpha) of single cells exposed to nalidixic acid as a function of a) time, and b) length. In this graph, only data points between 1.5 to 3 hr have been used. Black dots are single points from 48cells, red squares are binned data, and red line is a linear regression to the binned data. Error bars are standard deviation.

replication is continuous, indicating that exposing cells to cephalexin has no effect on their growth rate. Figure 46a displays the same data as in Figure 43b, indicating that the nalidixic acid affected the cells between times 1.5 hr and 3 hr. The growth rate of cells clearly decreases during this time, and after the removal of the antibiotic, the cells revert to their high growth rate as before the exposure. To only consider data where the cells were under the effect of nalidixic acid, we removed all points before 1.5 hr and after 3 hr and plotted the growth rate as a function of length in Figure 46b. The negative slope verifies that an increase in length results in a decrease in growth rate. We predict that the intercept of this line will reflect the cells' growth rate constant DNA concentration. Figure 47 compares the fits for cells exposed to both antibiotics. Remarkably, the analysis of the growth rate of cells exposed to cephalexin, where DNA concentration is constant (Figure 45), reveals a horizontal line with a value very close to the intercept of the red line.

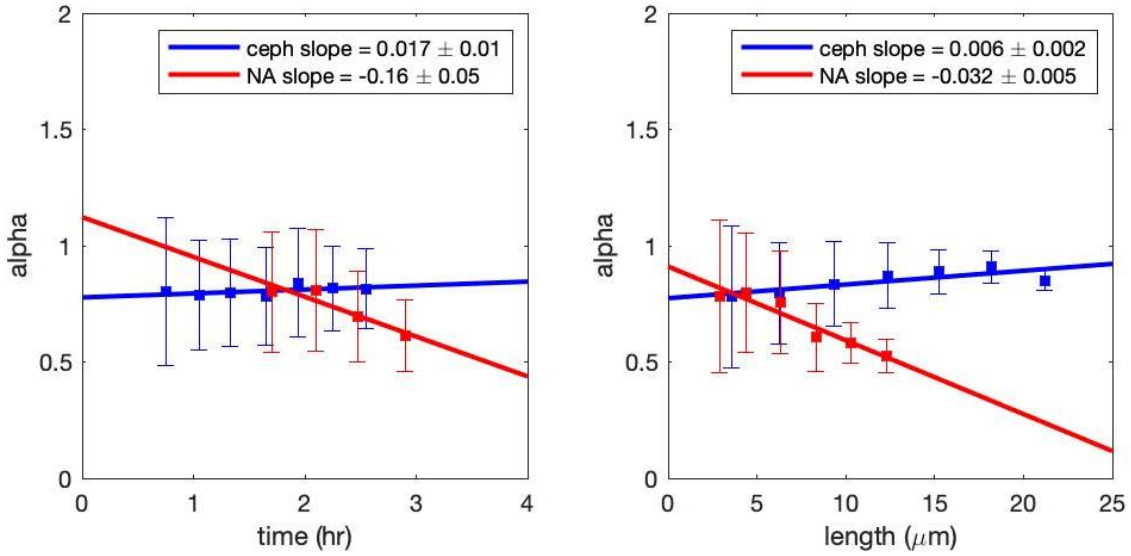


Figure 47 A comparison between fitted data found for cells exposed to cephalexin (Figure 45) and nalidixic acid (Figure 46). Blue lines indicate how growth rate of cells with continuous DNA synthesis does not change with respect to either time or length, and red lines reveal a drastic decrease in growth rate of cells with inhibition of DNA replication with respect to both time and length.

To further verify this hypothesis, we have measured the DNA concentration in each cell as it filaments (see Section 4.1.3 for DNA measurement details), as well as a constitutively expressed fluorescent protein as a representative of all cellular proteins (see Section 4.1.1). As expected, the method of filamentation does not affect protein concentration. This value remains constant on average and has no dependency on either time or length (Figure 48c-d, Figure 49c-d).

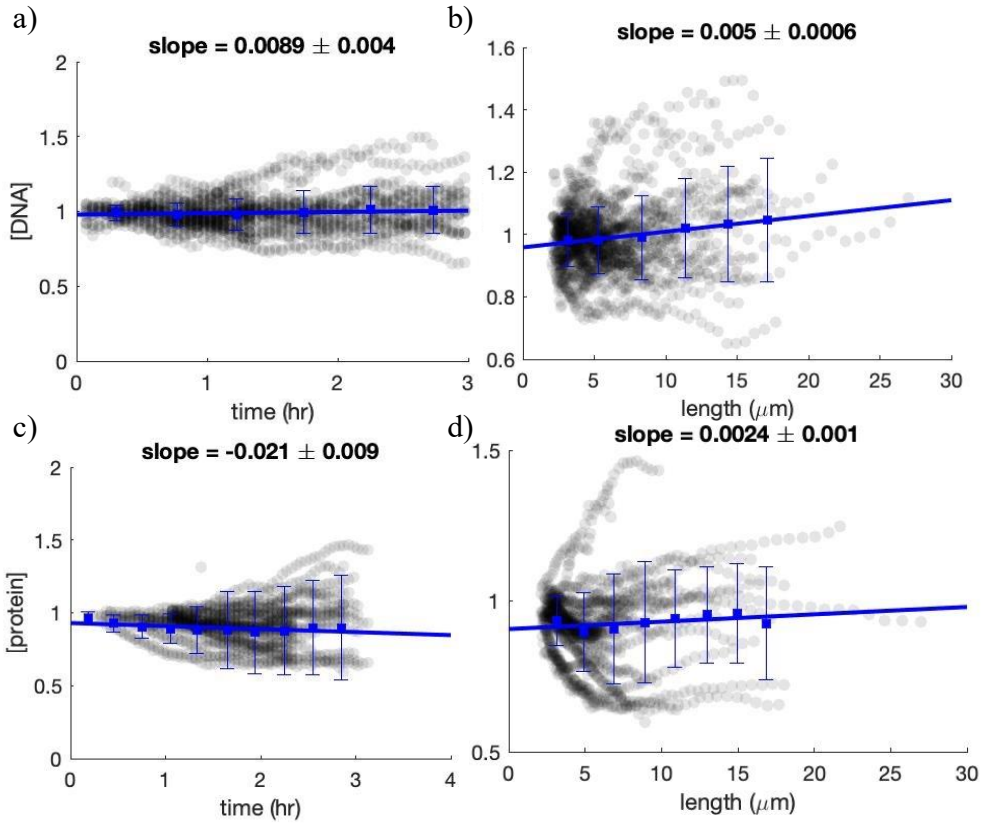


Figure 48 It has been previously shown that in cells exposed to cephalexin, DNA replication continues as the cell elongates[102]. Our results verify this, and also indicate the validity of our DNA concentration measurement method. Concentration of DNA as a function of a) time, and b) length. Concentration of red protein expressed from a plasmid as a function of a) time, and b) length. Black circles are values calculated for 50 single cells, blue squares are binned data, and blue line is a linear fit to binned data. Error bars are standard deviation.

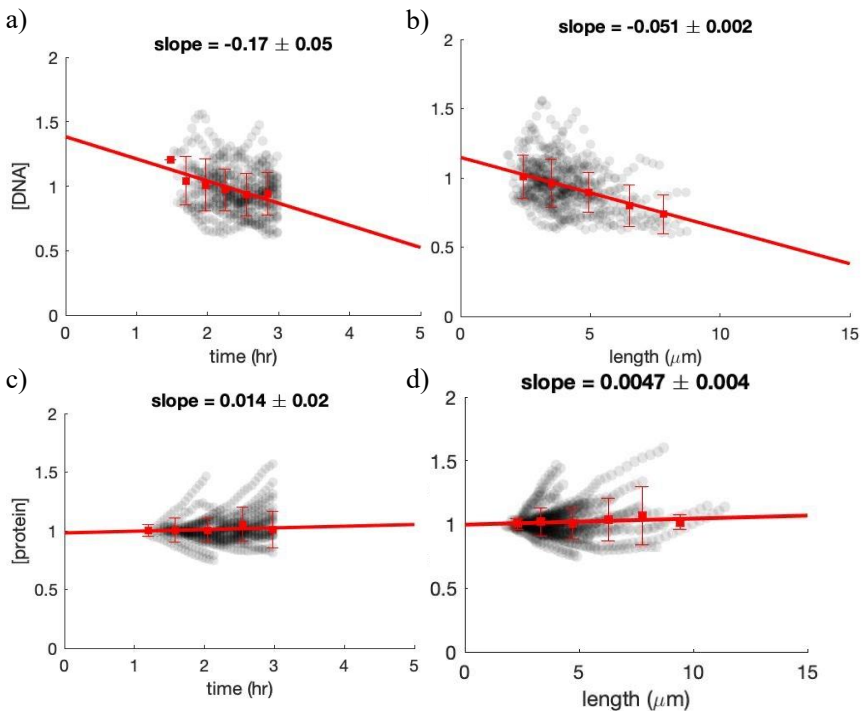


Figure 49 Nalidixic acid has previously shown to inhibit DNA replication in cells while also not allowing them to divide[101]. Our results presented here verify this, and also indicate the validity of our DNA concentration measurement method. Concentration of DNA as a function of a) time, and b) length. Concentration of red protein expressed from a plasmid as a function of a) time, and b) length. Black circles are values calculated for 48 single cells, red squares are binned data, and red line is a linear fit to binned data. Error bars are standard deviation.

DNA concentration, on the other hand, is expected to be constant for cells exposed to cephalexin, and Figure 48a-b point towards this, implying the continuation of DNA synthesis during elongation. In addition, DNA concentration is expected to decrease for cells exposed to nalidixic acid. Figure 49a-b show that the DNA concentration is indeed decreasing with respect to time and length. This implies that as a cell elongates, DNA is either not synthesized or synthesized at a lower rate compared to other material in the cell, causing the ratio of DNA to other biomaterials to continuously decline. Figure 50 compares the results of cephalexin and nalidixic acid and confirms a significant decrease in DNA concentration for cells exposed to nalidixic acid compared

to cells exposed to cephalexin. This figure also validates the accuracy of our DNA concentration measurement method.

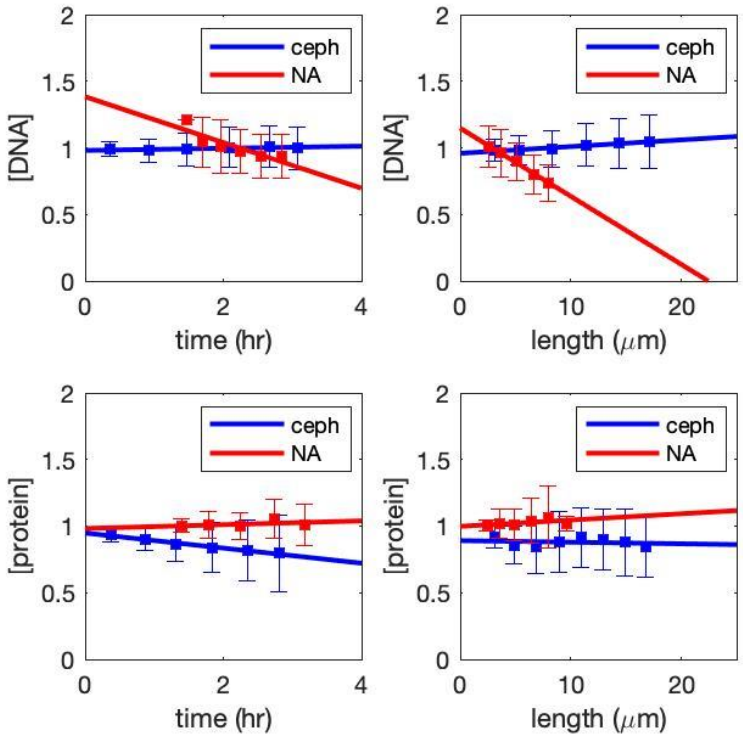


Figure 50 A comparison between the results of DNA and protein concentration found in cephalexin (Figure 48) and nalidixic acid (Figure 49). Blue lines indicate how DNA and protein are being synthesized at a constant rate throughout the entire experiment, including when cells were exposed to cephalexin. Red lines indicate that protein synthesis remains constant at all times including when cells were exposed to nalidixic acid, however DNA concentration decreases, indicating inhibition of DNA synthesis.

Next, we examined the correlation between the DNA concentration and growth rate. The results we present confirm that the growth rate of the cell in both cases is linearly proportional to the concentration of DNA, and an increase in growth rate is observed when the DNA concentration increases (Figure 51). When the binned data achieved from nalidixic acid experiments and cephalexin experiments were plotted on the same graph, one line could fit all data points well as can be seen in Figure 52.

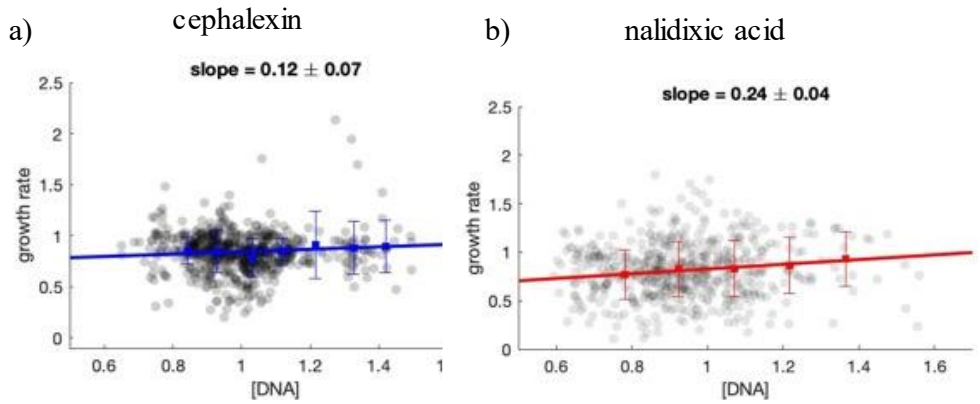


Figure 51 Correlation between growth rate and DNA concentration of cells exposed to a) cephalexin, where DNA concentration is not changing much for each cell in each cycle, and b) nalidixic acid, where DNA concentration is decreasing with an increase in length for each cell. Positive slope in both graphs varifies a positive correlation between these two parameters. Black dots are data from single cells, blue squares are binned data, blue line is a linear regression to the binned data, and error bars are standard deviation.

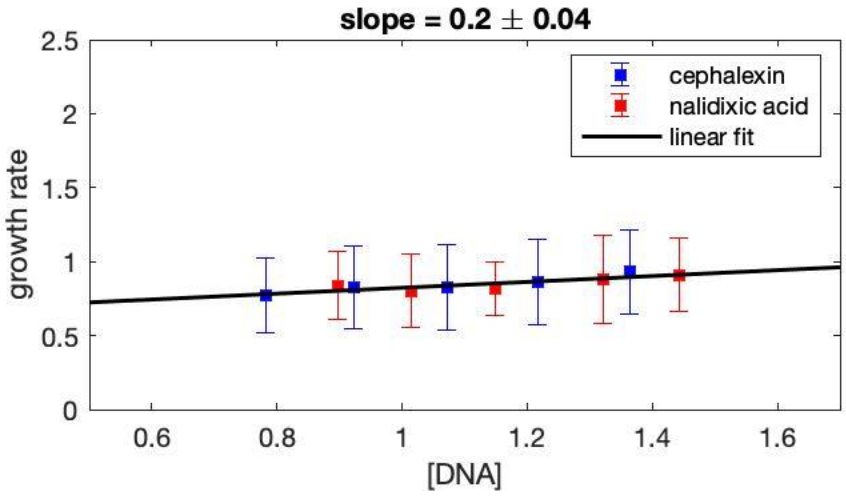


Figure 52 Same data as in Figure 51. Here, only the binned data are represented and all data are fitted to one line (black line).

4.7 Inhibition of cell division by overexpression of *sulA* confirms the results of division inhibition by cephalexin

Since antibiotics might have other effects on cellular growth besides inhibiting DNA replication or blocking cell division, we induced filamentation and carried out the measurements described in Section 4.6 without the use of antibiotics. Instead, we regulated the expression of the division inhibitor protein SulA, which allows DNA replication but prevents cell division by inhibition of FtsZ polymerization. This strain has been shown previously to grow at constant rate as in the case of growth under cephalexin in our experiments[107]. To regulate the expression of the *sulA* gene, we used an IPTG inducible plasmid that expresses SulA and GFP-Fis under the lac promoter (see Section 4.1.1). Figure 53 shows a sample channel of the mother machine containing a mother cell carrying this plasmid. IPTG is added after 15 minutes to the medium feeding this channel, and it can be seen that GFP becomes visible around 15 minutes after that. The cell also stops dividing after one hour and elongates until 15 minutes after the IPTG is removed (IPTG is removed at $t = 2.45hr$), however, it takes much longer for the GFP signal to disappear. The delay witnessed in this type of experiment could be due to our chosen method of filamentation. After addition of IPTG, it takes time for the amount of sulA in the cell to be enough for inhibiting division, and after removal of IPTG, FtsZ proteins need time to polymerize and create the septum for the cell to divide. Figure 54 shows how cells undergo one or two divisions before starting to filament. It can also be seen that the cells continue their exponential growth, as observed in the experiments with cephalexin (Figure 43a). This is expected since DNA replication continues at a constant rate during cell elongation under excess SulA.

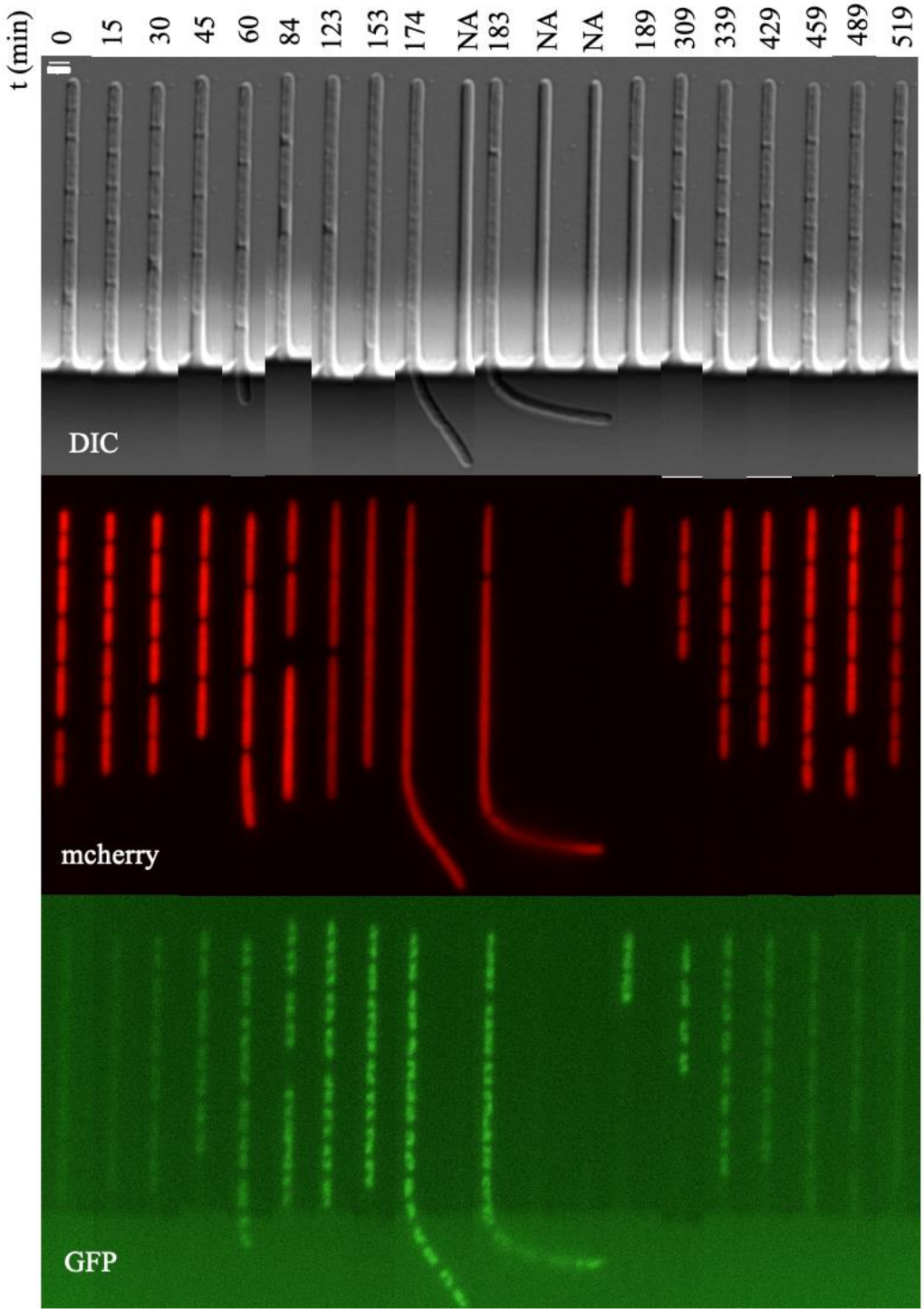


Figure 53 Time-lapse of one sample channel when induced with IPTG in three channels. IPTG was added at 15 min and removed at 165 min. White line on the top left is 2 μm.

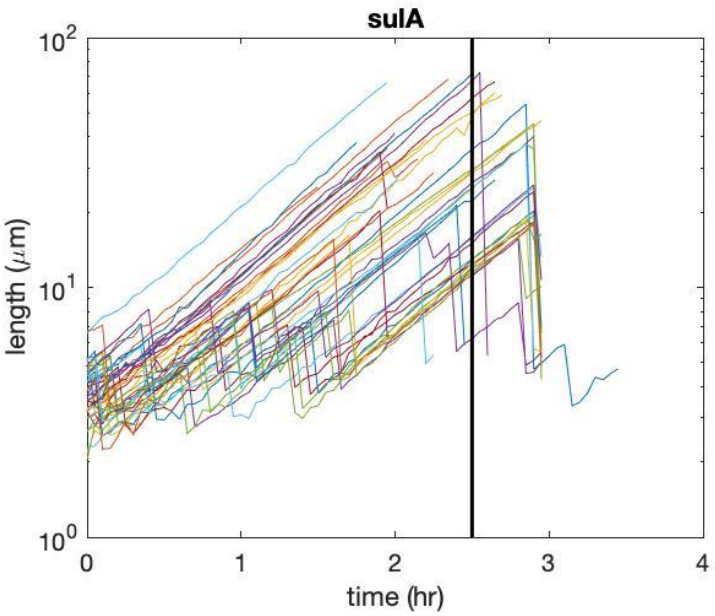


Figure 54 Growth of single cells carrying a plasmid that expresses excess sulA when induced with IPTG. In this experiment, IPTG was added at $t = 0$. Some cells undergoe few divisions before they start to filament. The growth of cells seems to remain exponential even after induction of excess sulA. The IPTG was removed at time $t = 2.5hr$ indicated by a vertical black line, and cells begin to divide again shortly after that.

Analysis similar to the previous section on single cells exposed to excess sulA confirm our hypothesis of how growth rate is unaffected when DNA concentration is constant. Figure 55 shows that the growth rate of single cells does not change as a function of time or length. Figure 56 reveal that neither the DNA nor the protein concentration changes during such experiments as well. Note that the slight increase in Figure 56a is due to the fact that IPTG induces GFP expression, and with time this expression slightly increases before it can stabilize. In addition, a positive slope between the variables in Figure 57 points to a positive correlation between the growth rate and the DNA concentration. Similar analysis with another plasmid that constrains DNA synthesis (such as a plasmid that expresses excess dnaA) could be performed in a future study to strengthen the results obtained from experiments with nalidixic acid.

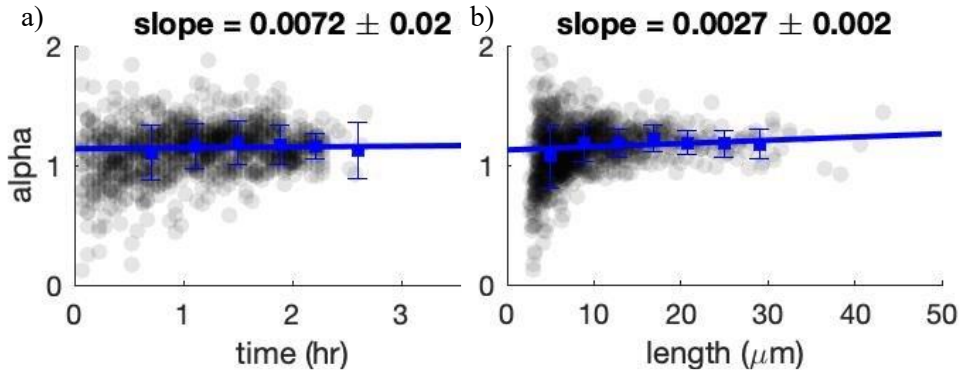


Figure 55 Growth rate (alpha) for 40 single cells exposed to excess sulA as a function of a) time, and b) length. Black dots are measurements from single cells, blue squares are binned data, and blue line is a fit to the binned data points. Error bars are standard deviation.

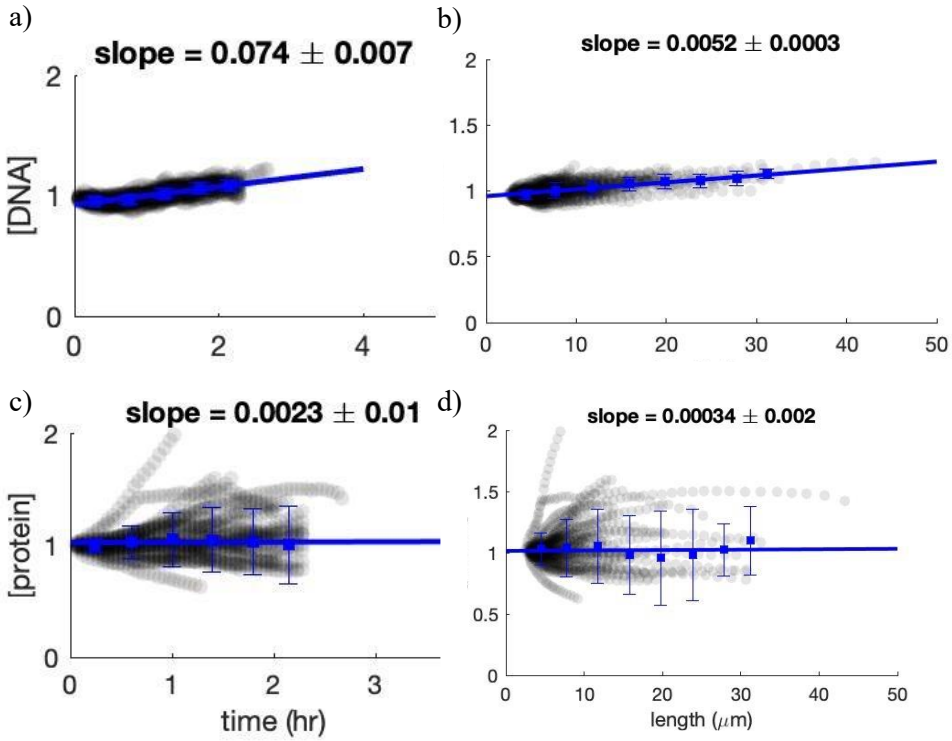


Figure 56 Concentration of DNA as a function of a) time, and b) length. [DNA] remains constant during the entire experiment. Concentration of protein as a function of c) time, and d) length, which remains constant as well. Black dots are measurements from single cells, blue squares are binned data, and blue line is a fit to the binned data points. Error bars are standard deviation.

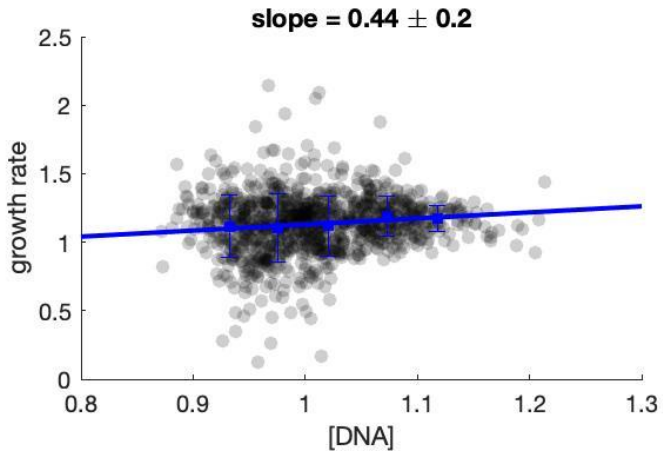


Figure 57 Correlation between growth rate and DNA concentration for cells with excess sulA, where DNA concentration is not changing much for each cell in each cycle. Positive slope varifies a positive correlation between these two parameters. Black dots are data from single cells, blue squares are binned data, blue line is a linear regression to the binned data, and error bars are standard deviation.

5.0 Aging as a consequence of ecosystem-level bacterial adaptability

When encountering an unfavorable environment, (micro)organisms can have one or more of the following three behavioral responses: they can migrate away from the new environmet[108], they can adapt to it[109], or they die[110]. Migration away from harmful environments can be propelled by chemical gradients (chemotaxis[111]), thermal gradients (thermotaxis[112]), light gradients (phototaxis[113]), and/or ionic gradients (electrotaxis[114]). While our understanding of the various modes of taxes has been greatly advanced, the modes of adaptation to new environments are still largely unclear[115].

The process of random gene mutation enables the co-existence of genetically diverse cells of a bacterial species upon which natural selection acts. A favorable outcome of this process is the expansion of cells with mutations in genes that favor survival and proliferation in the new environment. However, an unavoidable 'side effect' of this process is the emergence of many deleterious gene variants in cells. In its extreme form, this can lead to cell death. In other cases, the presence of deleterious mutations does not manifest until late in the life cycle of a cell. This in turn would allow mutant cells to grow and proliferate similar to their wild-type variant, which will lead to the preservation of the mutation in the population. An example of such mutation in humans is the Huntington's disease, whose effects start in people at ages over 30 or 40. In such cases, the growth of the population might seem normal and almost identical to the wild-type population. Yet, the consequence of non-lethal random deleterious mutations on the physiology and function of individual cells, and on the cell ecosystem as a whole, is not well understood. In this chapter, we focus on the details of how some gene deletions that appear to have almost no consequences to the population proliferation despite the cell's limitation in key metabolic processes, affect cellular

growth at the single-cell level. We focus on the long-term fate of cells with a single gene deletion in nutrients rich and poor environments. Specifically, we investigate the mode by which cellular homeostasis is maintained in the presence of metabolic gene inactivating mutations by observing population growth of multiple strains in microbatch cultures and single cells in homogeneous environments. Our single cell experiments provide new insights into bacterial senescence and the bacterium's various modes of death and identify cellular characteristics of aging.

5.1 Materials and methods

5.1.1 Bacterial strains, growth conditions and reagents

E. coli K12 BW25113 and its isogenic mutant derivatives were obtained from the Keio *E. coli* mutant collection[81,82]. A complete list of all used strains is presented in Appendix D. All mutants harbored the single copy number plasmid pXX563 containing kanamycin resistance[54]. The cells used in the mother machine experiments carried two additional plasmids, the chloramphenicol-resistant medium copy number plasmid pZA under the control of the constitutive viral $\lambda - P_R$ promoter[78], and the plasmid PRJ2001-GFP-Fis which was used for chromosome labeling (a kind gift from John F. Marko[116]).Three different growth media were used in the experiments: Luria Broth (LB, used as a rich medium for rapid growth), M9 minimal medium supplemented with 1 g/L casamino acids and 4 g/L glucose (M9CG, used for medium growth), and M9 minimal medium supplemented with 4 g/L glucose (M9G, used for slow growth). Depending on the identity of the strain being used either 25 $\mu g/ml$ kanamycin, or 30 $\mu g/ml$ chloramphenicol, or both were added to the medium.

5.1.2 Microbatch culture experiments

All isogenic single gene deletion mutants and *wt E. coli* strains were stored in microtiter plates at -80℃. On the day before the experiment, strains were picked from frozen stocks and grown overnight at 37°C with strong agitation (240 rpm) in a selected medium and with the appropriate antibiotic. The overnight cultures were then further diluted to optical density (OD_{600}) ~ 0.001 in fresh medium. 250 µl of each cell type was then transferred into a 96-well cell culture plate in triplicates and grown up to 16 hours at 37℃ in a plate reader (Infinite 200, Tecan Trading AG, Switzerland). The growth of the various cell-type populations in the different growth conditions was characterized by obtaining optical density measurements at 600nm every 5 minutes with shaking in between measurements. All experiments were repeated three times.

5.1.3 Microbatch culture data analysis

In a previous study, growth of non-lethal *E. coli* isogenic single-gene knockout mutants (the Keio collection) on semi solid agar was investigated, and three *E. coli* colony growth values – lag time of growth (LTG), maximum growth rate (MGR) and saturation point of growth (SPG) – were measured using a high-throughput measurement system. The results were able to identify a subset of genes whose deletion delayed colony formation or changed saturation point and presented a better understanding of the range of phenotypes that exist in this strain[54]. We chose 65 mutants among this subset which are listed in Appendix D.

Specific growth rate of bacterial population at time t_i was calculated using the equation: $\mu_i = ln\left(\frac{N_{i-1}}{N_i}\right)/(t_i - t_{i-1})$, where N_i is the population concentration (OD_{600}) at time t_i (Figure

64). MGR and SPG were found for each mutant from their growth curves as follows (Figure 58 and Figure 59). First, all OD_{600} values that were below zero were set to zero, since a negative value for OD does not have any meaning and is a result of a measurement error. Then the background measurements, obtained in each experiment from a blank sample, where the well contained only the growth medium, were subtracted point by point from each population growth graph. To find MGR, the maximum of the first derivative of $ln\left(\frac{N}{N_0}\right)$ was found, and that point plus five points before and five points after were fitted to a line and the slope of that line was calculated as the MGR. For SPG, the last 20-50 points of the graph were forced to fit to a horizontal line, and the value of that line was calculated. Figure 58 illustrates in more detail how these values were extracted for one sample graph (wild type in LB). Figure 59 presents the growth curves of microbatch cultures of all 65 mutants in three media, and the fitted lines for MGR and SPG. The values extracted from these graphs are represented in Figure 65.

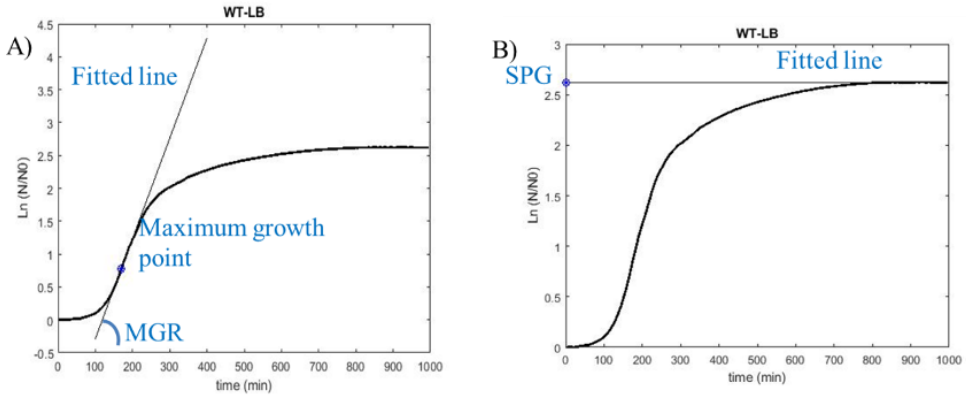


Figure 58 Extraction of growth parameters in microbatch cultures. (A) Maximum Growth Rate (MGR): maximum of the first derivative of *ln(N/N0)* (where *N* is OD_{600} and N0 is the starting OD_{600} values) was found and 5 points before and 5 points after were fitted to a line. MGR is the slope of that line. For wt *E. coli* in LB, the MGR was found to be 0.15 min^{-1}. B) Saturation Point of Growth (SPG): the last 20-50 points on the growth curves were fitted to a straight line and the value of *ln(N/N0)* was taken as the SPG. For wt in LB, SPG was found to be 2.61.

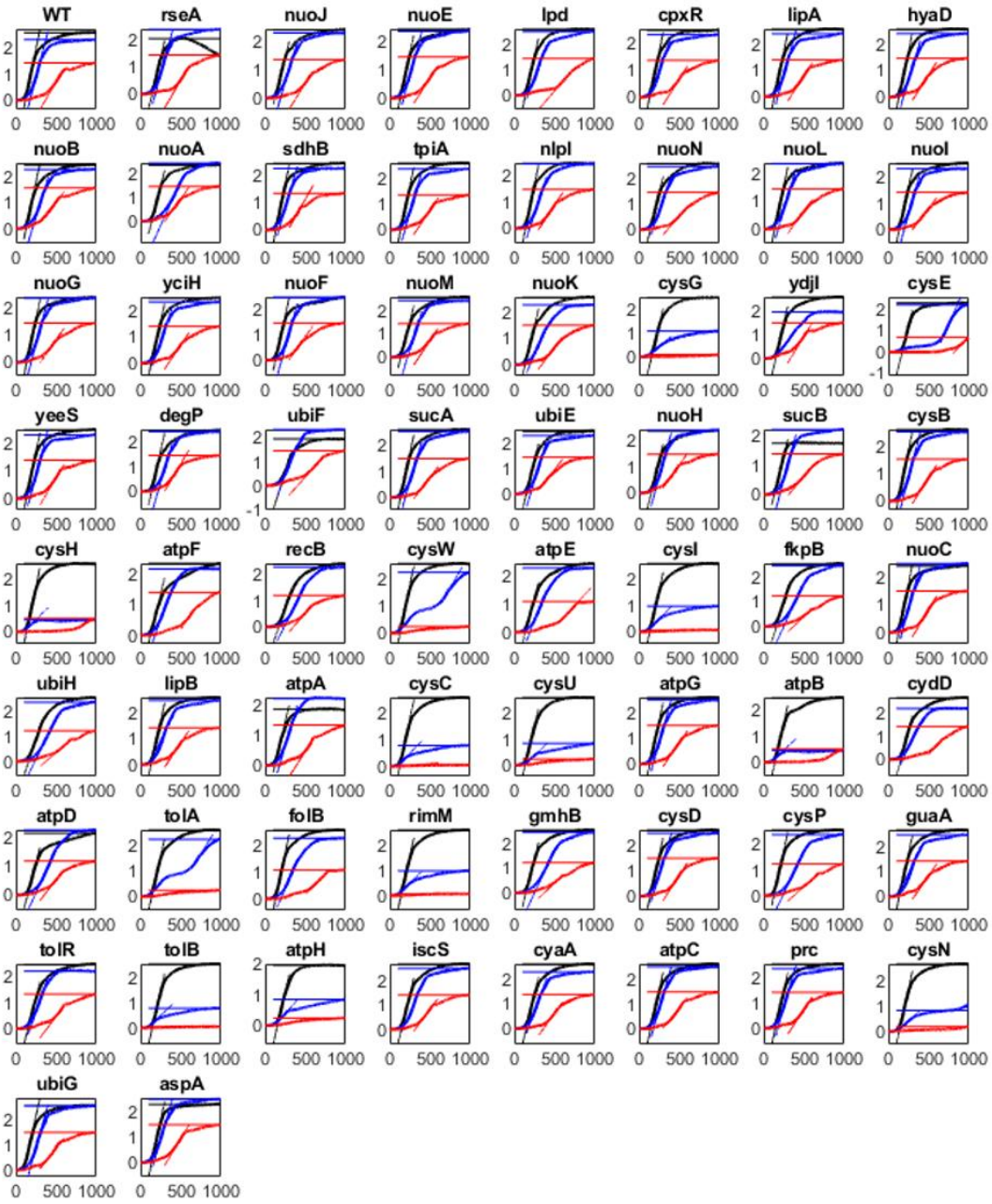


Figure 59 Population growth using microbatch cultures for all 65 mutants and wild-type in LB, M9CG, and M9-G growth media. Straight lines indicate fit lines for extracting parameters as shown in Figure 58. Horizontal axis is time (min) and vertical axis is ln(N/N$_0$) where N is OD$_{600}$ and N$_0$ is the starting OD$_{600}$ values. Black lines represent the medium LB, blue lines M9CG, and red lines M9G. All negative values of OD$_{600}$ were set to zero. After reaching saturation, some of the growth curves show a decline which is the indication of the death phase. This phase was neglected and deleted from the curve for parameter extraction. To see extracted values see Figure 65.

5.1.4 Data reduction and clustering

Principal Component Analysis (PCA) was performed to extract the underlying dimensions of the dataset characterizing the microbatch cultures growth curves. All the parameters (MGRs and SPGs) distinguishing the different growth curves of all 65 mutants + 1 wild type strains grown in 3 different media form a 66×6 matrix, which we label X. The built-in MATLAB function "svd" (singular value decomposition) was used to decompose X as a product of three matrices: $X = U\Sigma V^*$, Where U (66×66) contains the left eigenvectors and V (6×6) contains the right eigenvectors. The matrix Σ (66×6) contains the eigenvalues, or the principal components, along its main diagonal. The scree plots are shown in Figure 60.

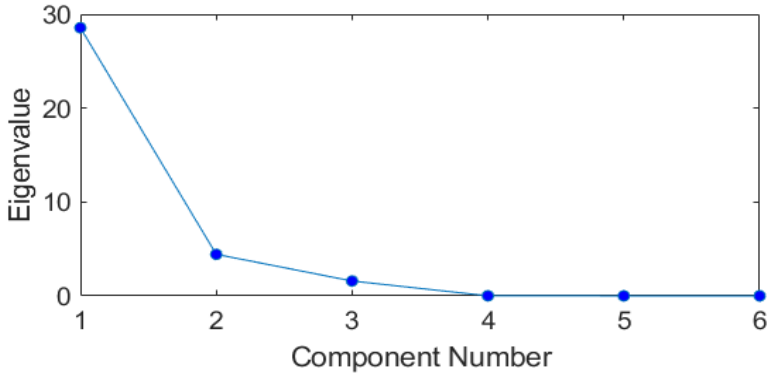


Figure 60 PCA scree plots of six growth parameters. Calculating the cumulative sum of the eigenvalues indicates that the first three are enough to explain 99% of the variance in the data.

The cumulative sum of the eigenvalues indicated that more than 99% of the variance can be explained by the first three principal components, and so it is reasonable to reduce the dimension of these measurements from six to three. In addition, the Pearson correlation coefficients between the MGRs and SPGs were calculated (Figure 61). To reduce the dimension, the first three rows from the right side of matrix V were deleted and multiplied by X from the right side: $XV_r = U_r\Sigma_r$.

This is now a representation of the projection of the entire data on three principal components, which are presented in Figure 62.

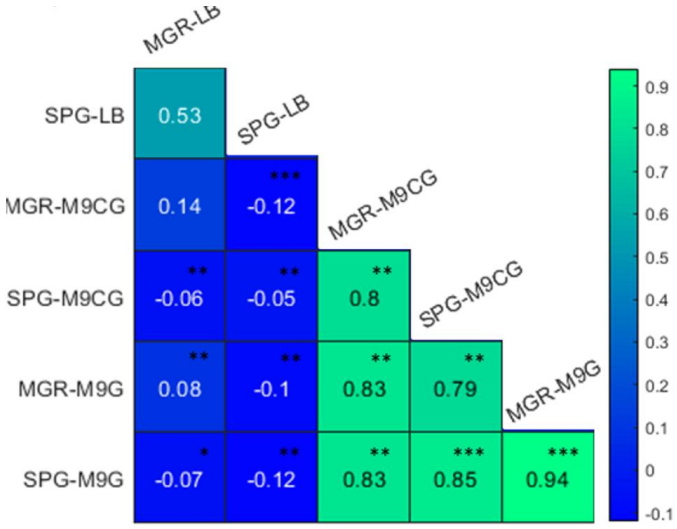


Figure 61 Pearson correlation coefficient among set of parameters found in three different environments (LB, M9CG, and M9G media) for 65 *E. coli* single gene deletion mutants and the wild type from their minibatch growth phenotypes. The correlations are high and significant among most parameters. *$p = 0.01$, **$p = 0.001$, ***$p = 0.0001$.

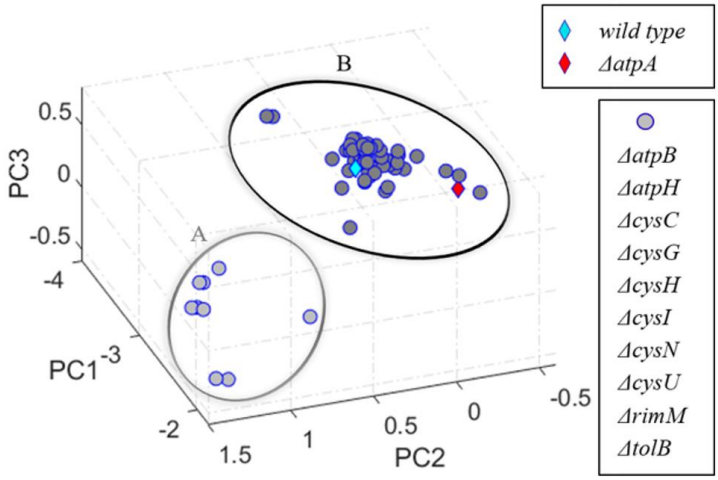


Figure 62 Mutants are categorized into two clusters (A and B) based on three calculated principle components. Cluster A includes mutants that exhibit very slow growth or low saturation under low nutrition conditions (M9CG and M9G). Both wild type and ΔatpA are in cluster B, illustrating similarities in their population growth.

Clustering of data was performed using the hierarchical method. In this method, first the similarities between every pair of mutants are found by measuring the distance between them, then using that information, each pair is linked into a binary cluster. In the next step these binary clusters are linked again to form larger clusters. This process is repeated until the number of clusters reaches the intended value. We used the built-in MATLAB function "clusterdata" for this purpose and set the maximum number of clusters to two.

5.1.5 Oxygen consumption rate and extracellular acidification rate

Oxygen consumption rate (OCR) and extracellular acidification rate (EACR) were measured using a Seahorse assay modified for bacterial cells[100]. The wells of a 96- tissue well microplate (Agilent) were coated by poly-L-lysine (0.1% (w/v) Poly-L-lysine from Sigma-Aldrich) in order to bind and immobilize the bacteria in preparation for the measurements. This was achieved by adding 15 μl of 0.0004% and leaving the plate in a hood overnight to allow the evaporation of the solution. The following day, each well was rinsed with 200 μl of dH_2O to remove excess poly-L-lysine molecules. This has been shown to have a negligible effect on the metabolism of *E. coli*[118].

Wild type and Δ*atpA* cells were grown in LB overnight in an incubator at 37℃ with shaking at 240 rpm. The following day, they were diluted 100x in the three different media (LB, M9CG, and M9G) and regrown at 37℃ while shaking at 240 rpm until $OD_{600} \sim 0.3$. The cultures were then diluted further to $OD_{600} \sim 0.02$ in the associated fresh medium. 90 μl of each culture was added to the poly-L-lysine coated wells of the microplate in triplicate wells, and the plate was centrifuged for 10 minutes at 4,000 rpm. An additional 90 μl of the associated fresh medium was then added to each well. In addition to the bacterial cultures, three wells were filled with clear

media (no cells) only for control, and the plate was incubated for 1 hour before placing it in an Agilent Seahorse XFe96 Analyzer for measurements. The OCR and EACR measurements of the control wells are represented in Figure 63. For normalization, multiple bright field images of multiple wells were taken, and an approximate concentration of cells was calculated using imageJ. Examples of these images can be seen in Figure 63.

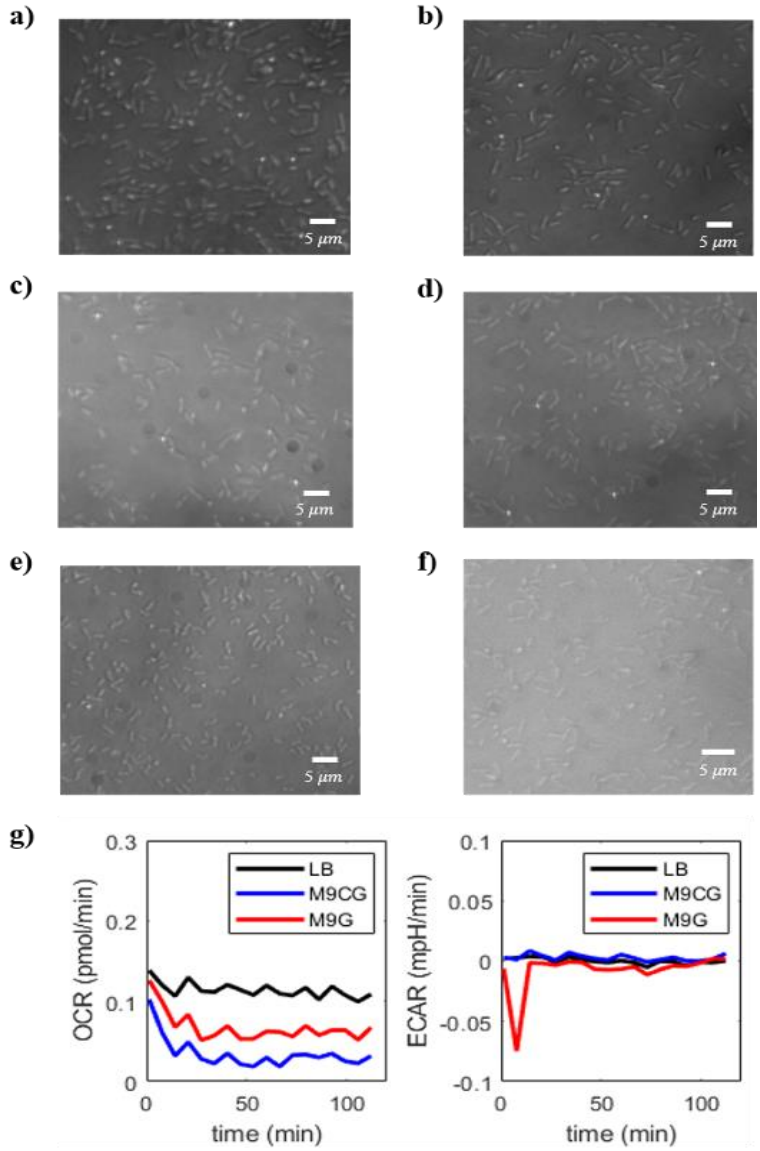


Figure 63 Examples of brightfield images taken from a microplate prepared the same way as used in the seahorse experiment. a) wild type in LB, concentration was found to be 61±30 cells/mm. b) Δ*atpA* cells in LB, concentration was found to be 50±47 cells/mm. c) Wild type in M9CG, concentration was found to be 70±35

cells/mm. d) Δ*atpA* cells in M9CG, concentration was found to be 68±22 cells/mm. e) wild type in M9G, concentration was found to be 73±18 cells/mm. f) Δ*atpA* cells in M9G, concentration was found to be 77±27 cells/mm. g) Background measurements for oxygen consumption rate and extracellular acidification rate for samples where there are no cells present in the liquid.

5.1.6 Single-cell experiments in mother machine

Cell length and protein expression were measured using the microfluidic trapping device, the "mother machine"[49] (for details on fabrication of mother machine see Chapter 2.0). The wild type strain containing the plasmids PZA3R-mcherry and PRJ2001-GFP-Fis, or the Δ*atpA* strain with these two plasmids and also pXX563, were grown from an agar plate overnight at 37°C with strong agitation (240 rpm) in the selected medium (LB, M9CG, or M9G) and with appropriate antibiotics. The overnight cultures were then diluted 100-fold in fresh medium and grown at 32°C until early exponential phase, optical density (OD_{600}) of 0.1-0.2. Cells were then concentrated into fresh medium to an OD_{600} ~ 0.3 and loaded into the trapping device that was then mounted on a Zeiss Axio Observer microscope with a 100× objective. Temperature was maintained at ~ 30°C using an in-house made incubator. Fresh medium was flown into the device at a rate of 1 ml/hr throughout the entire experiment. All media getting pumped into the device contained 1mM Isopropyl β-D-1-thiogalactopyranoside (IPTG**)** for GFP-Fis induction. In some experiments, propidium iodide (IP) from LIVE/DEAD ™ *Bac*Light ™ Bacterial Viability Kit (ThermoFisher Scientific) was added to the medium to identify dead cells. In such experiments, cells did not contain the plasmid PZA3R-mcherry, and so the red channel was used to detect the dye. Images of the channels were acquired every 3-6 minutes in phase contrast and fluorescence mode using a CCD camera (Zeiss AxioCam MRm). This resolution ensures a continuous measurement relative

to the typical timescales of change in both cell size and protein content, while minimizing the damage to the cells. The software TLM tracker[62], along with custom designed MATLAB codes were used to measure the length and fluorescence intensity of the mother cell.

5.2 *E. coli* single gene deletion mutants display media-dependent growth in microbatch cultures

Random gene inactivation in cells of bacterial colonies can have substantial consequences for bacterial physiology in an environment-dependent fashion. To explore this idea further, first we tested, in microbatch cultures, the growth of wild-type (*wt*) *E. coli* BW25113 and its 65 isogenic single gene deletion mutants (listed in Appendix D). We selected these mutants because in a previous study they displayed the most severe growth defects in a soft-agar based colony assay under single-substrate limited growth[54]. We used three different growth media as proxy for different environments for *wt E. coli* cells as mentioned in Section 5.1.1: a rich medium (LB) that results in rapid growth (25 minutes division time at 37°C); a medium resulting in average growth (M9CG, 50 minutes division time at 37°C); and a glucose-limited medium resulting in slow growth (M9G, 75 minutes division time at 37°C). In this microbatch plate setup, favored nutrients are consumed first by the growing *E. coli* population, which leads to a continuous change in substrate concentrations and the pH of the environment[120,121]. However, in contrast to that seen in soft-agar colony assays[114], the emergence of localized microenvironments is not expected to occur.

We observed significant differences in the growth curves (Figure 59) and derived growth rates (Figure 64) of *wt E. coli* and its isogenic single gene deletion mutant strains in the three

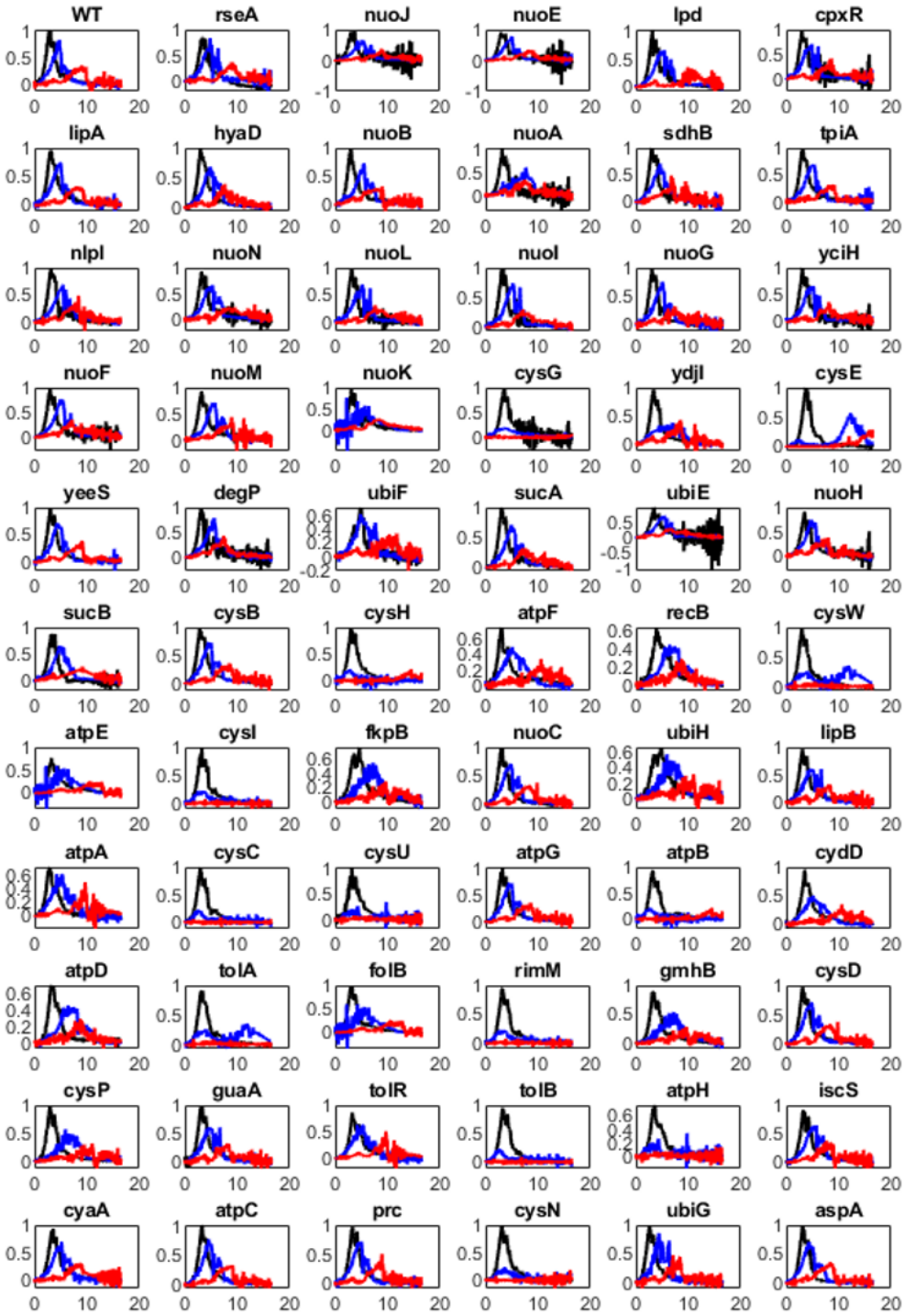


Figure 64 Specific growth rate of bacterial populations. The calculated specific growth rate (μ) for wild-type *E. coli* and its 65 tested isogenic mutants in LB, M9CG, and M9-G growth media are shown. The value for μ_i at time t_i is derived from the growth curves of Figure 59 using the equation, $\ln(N_{i-1}/N_i)/(t_i-t_{i-1})$, where N_i is the population concentration (OD_{600}) at time t_i. Labels are the same as in Figure 59.

different growth media. To quantitatively assess these differences, we extracted two parameters from the obtained growth curves: maximum growth rate (MGR) and saturation point of growth (SPG), using an established method[54] explained in Section 5.1.3 (Figure 58). We observed significant differences in the calculated MGR and SPG values for the *wt* and 65 single gene deletion mutant strains (Figure 65) in the three growth environments.

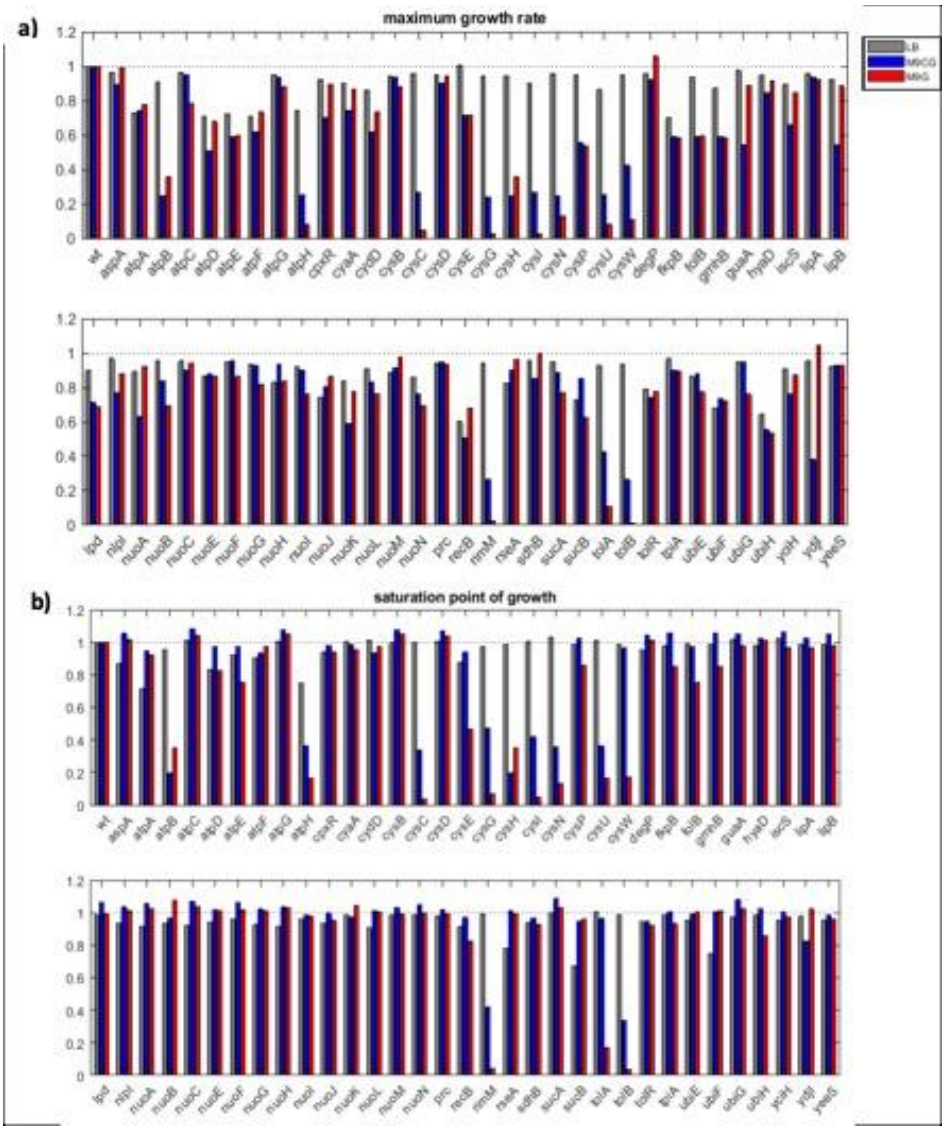


Figure 65 Histograms of MGR and SPG values for *E. coli* cultures. (A) The MGR and (B) SPG values for wild type (wt) *E. coli* and the indicated isogenic single gene deletion mutants are shown. Values are normalized such that all parameters for wt E. coli are equal to one.

We then assessed the relationships among the strains by calculating the Pearson correlation coefficient among their six derived growth parameters (MGR and SPG in LB, M9CG, and M9G). The observed strong correlations among several of them (Figure 61) suggested that our data could be reduced to fewer parameters. Thus, we performed principal component analysis (PCA) to obtain key attributes underlying growth features, and to identify mutants that display similar growth profiles. The principal components acquired from these analyses have indicated that 99% of the variability among the mutant strains can be characterized by using only three parameters (Figure 60). The projection of data points onto the principal components is shown in Figure 62. Subsequent hierarchical clustering grouped the strains into two main clusters. Mutant strains in Cluster A (depicted in light gray in Figure 62) exhibit significant population growth differences with *wt E. coli*. The remaining mutants in the second cluster (depicted in blue in Cluster B) display less difference from *wt E. coli* and less divergence in their growth profiles.

When grown in LB medium, some mutants with a deleted ATP synthase subunit (*ΔatpA, ΔatpD*, *ΔatpE, ΔatpF, ΔatpH*), a DNA repair and stress response related gene (*ΔrecB*), ubiquinone biosynthesis related genes (Δ*ubiF*, Δ*ubiH*), a gene related to electron transport (*ΔnuoJ*), a tricarboxylic acid cycle related gene (Δ*sucB*), a bacteriocin transport gene (*ΔtolR*), and one other strain (Δ*fkpB*), were found to have significantly lower growth rate compared to *wt*. However, of these strains, only five of them showed a significant lower saturation point compared to wt cells (*ΔatpA, ΔatpH, ΔrseA, ΔsucB, ΔubiF,*). When grown in M9CG or M9G media, a different set of mutants showed significantly lower growth rates. These strains consist of a mutant where a different ATP synthase subunit has been deleted (*ΔatpB*) and a mutant that also existed in the first set (*ΔatpH*), mutants with deleted genes related to sulfur metabolism (*ΔcysC, ΔcysG, ΔcysH, ΔcysI, ΔcysN, ΔcysU, ΔcysW*), mutants with deleted bacteriocin transport genes (*ΔtolA, ΔtolB*), and one

other mutant (*ΔrimM*). All these strains, except for two (*ΔcysW, ΔtolA,*), demonstrated a lower saturation point than wt *E. coli* cells in M9CG. Also, all mutants in addition to one more strain (*ΔcysE*) exhibited a lower SPG when grown in M9G. One strain had lower growth rate than *wt* in M9CG, but higher in M9G (*ΔydjI*). As expected, all values for growth rate are lower than wild type (except for the two cases of *ΔdegP* and *ΔydjI* in M9G), illustrating that no single gene deletion can increase the *E. coli* cells' normal growth rate in these experimental conditions. It is also notable that some mutants have a closer growth rate value to *wt E. coli* in M9CG than in LB; for example, *ΔatpA* exhibits higher similarity in growth to *wt E. coli* when in M9CG, and even higher similarity when grown in the nutrient limited medium, M9G.

Among the *E. coli* mutant strains there were single gene deletion mutants of the eight subunit-encoding genes of ATP synthase enzyme complex (*ΔatpA*, *ΔatpB*, *ΔatpC*, *ΔatpD*, *ΔatpE*, *ΔatpF*, *ΔatpG*, *ΔatpH*). The structure of ATP synthase in *E. coli* consists of two main modules (Figure 66); The F_0 module is comprised of subunits a, b, and c, is embedded in the inner bacterial membrane, and is responsible for proton transfer. The other module, F_1, with subunits $\alpha, \beta, \gamma, \varepsilon,$ and δ is membrane extrinsic and carries out the proton gradient-driven ATP synthesis[115,122]. The growth characteristics of the strains lacking either the δ subunit (*ΔatpH*) or the a subunit (*ΔatpB*) (Figure 66) places them in cluster A (Figure 62) (The a-subunits constitute two aqueous channels that allow proton migration, while the δ-subunit is part of the peripheral stalk and is bound to one of the three α-subunits[123]). In contrast, in the absence of the b-, c-, α-, β-, γ-, or ε-subunits, cells grow more similarly to *wt E. coli* (Figure 66). Such differences were not unique to ATP synthase; for example, we observed similar dichotomies among the growth of single gene deletion mutants encoding enzymes of the cysteine metabolism pathway (Figure 67). In summary, these data imply

that environmental conditions alter the growth of ATP synthase subunit deficient bacteria, with various dependences on intact ATP synthase activity.

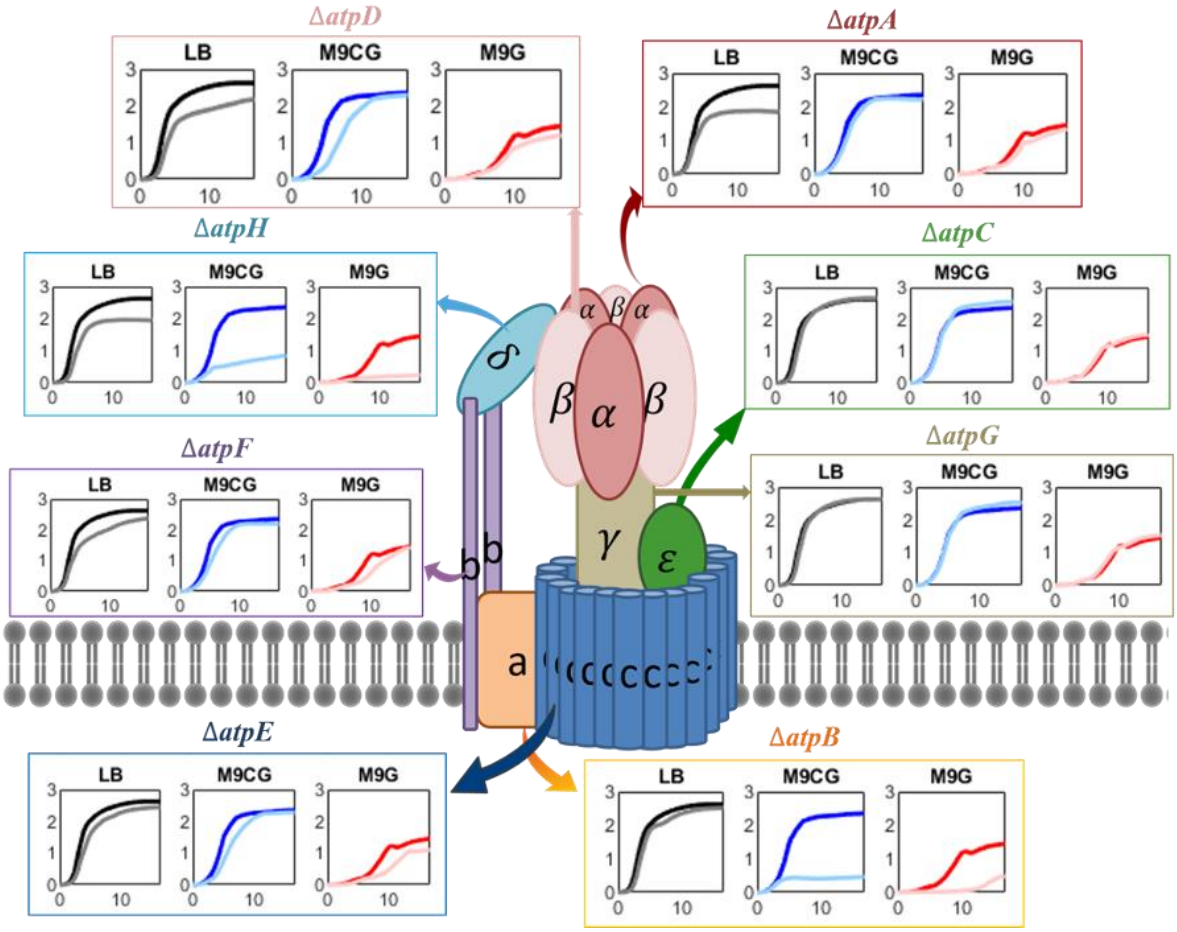


Figure 66 Growth similarities of E. coli ATP synthase subunit gene deletion mutants. The multi-unit enzyme complex, ATP synthase is embedded in the bacterial inner cell membrane and consists of two main modules; F0 contains subunits a, b, and c, while F1 contains subunits α, β, γ, ε, and δ. The graphs plotted for the growth of wild type (darker lines) compared to growth of indicated isogenic mutant strains (lighter lines) are shown. The x axis of all graphs is time (hr) and y axis is ln(N/N_0), where N is the concentration (OD_{600}) and N_0 is the initial concentration of the batch culture.

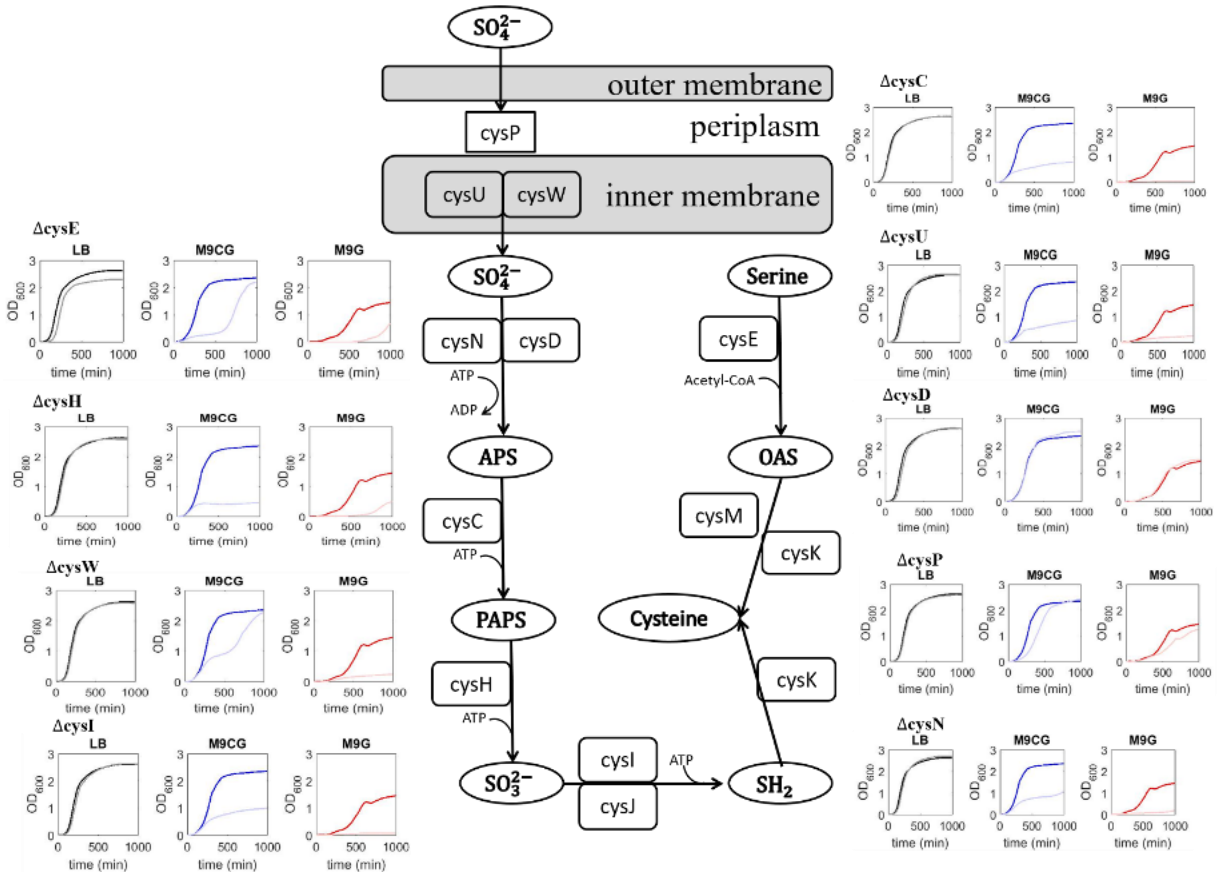


Figure 67 Cysteine metabolism pathway in E. coli bacteria . The population growth of both wild type (darker lines) and the single gene subunit deletion mutant (lighter lines) are shown in graphs around the pathway. Five of these mutants (Δ*cysC*, Δ*cysH*, Δ*cysI* ,Δ*cysN*, Δ*cysU*) are in cluster A from Figure 62, indicating slow or no growth in nutrient limited conditions (M9CG and M9G). The other four mutants (Δ*cysD*, Δ*cysP*, Δ*cysE*, Δ*cysW*) are in cluster B, indicating adaptive behaviors in all tested environments.

When a gene is randomly inactivated, *E. coli* cells likely undergo changes in their internal state in order to compensate for the loss of that gene's function. Consequently, the similarities among the observed growth profiles may reflect shared adaptive strategies. Although previous studies have investigated the roles of the residues of the ATP enzyme in detail[124,125], their specific roles in bacterial adaptation to environmental challenges remain unclear. In the following sections, we focus on the adaptive strategies of one such single gene deleted mutant.

5.3 Deletion of the ATP synthase *atpA* subunit increases cellular respiration in *E. coli*

We next wished to determine how major cellular metabolic functions in *E. coli* change when one of the ATP synthase subunits is inactivated. To this end, we focused on the *atpA* gene, which encodes the α-subunit of the ATP synthase (Figure 66). This subunit is part of the catalytic head of the enzyme that phosphorylates ADP and is connected to the enzyme's central stalk. Three active sites are formed at the interfaces of the three α- and β-subunits[53]. It has been proposed that the β-subunit serves as the high affinity catalytic site responsible for binding ADP[80]. Thus, *E. coli* cells may be able to synthesize ATP to some extent even in the absence of ATP synthase's α-subunit. Indeed, the specific growth rate (μ) of wt and Δ*atpA* cells are different in the three different growth media. In LB medium, *wt* and Δ*atpA* cells have significantly different growth rates; however, when grown in less rich media (M9CG and M9G) their growth becomes more similar (Figure 68a). Consequently, the calculated MGR and SPG of both populations display similar values only in environments where cells grow slower (Figure 68b).

To further investigate this phenomenon, we explored the association of population growth with *E. coli* cells' respiration and glycolysis. To this end, we measured the oxygen consumption rate (OCR) and extracellular acidification rate (ECAR) (proxy for cellular respiration and glycolysis, respectively[55]) of *wt* and *ΔatpA* cells by using the Seahorse technology[126]. *ΔatpA* cells exhibited a higher OCR and ECAR in all media (Figure 68c). However, this difference was most striking in rich LB medium while the difference was minimal in nutrient poor M9G medium. Cumulatively, these data indicate that during slow growth, *ΔatpA* mutant cells are able to compensate for their suboptimal ATP synthase activity, at least in part by consuming oxygen and glucose at a higher-than-normal rate. However, this compensatory mechanism(s) is not sufficient to maintain maximal growth in rich medium for *ΔatpA* mutant cells. The mode by which cells use

oxygen or glucose at a higher rate in the absence of a certain gene products requires further investigation.

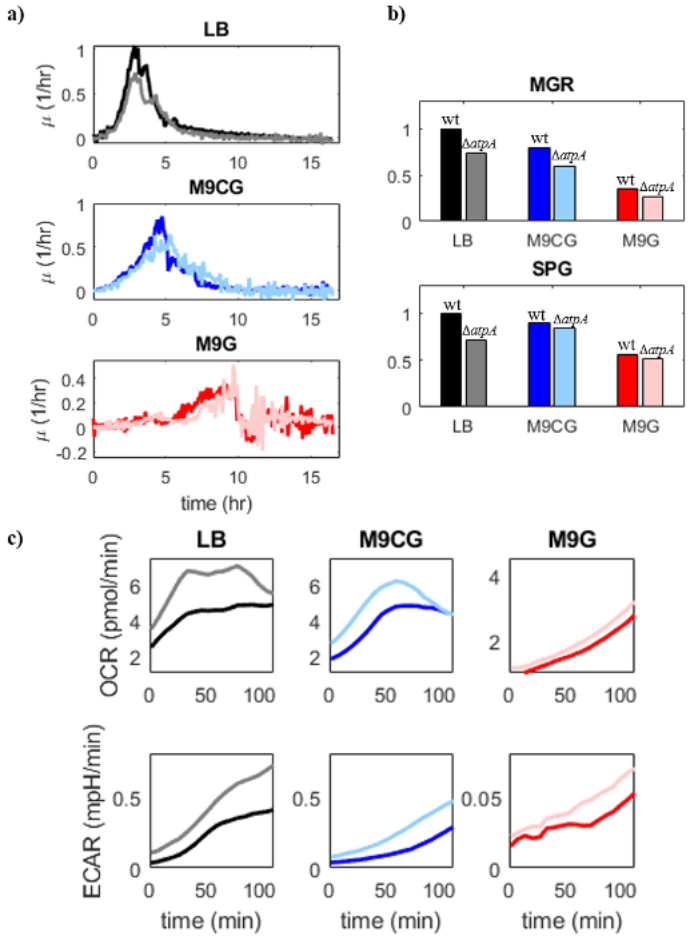


Figure 68 Differences in the physiology of Δ*atpA* from wt *E. coli* cells in varied media . Darker colors represent wild type and lighter colors represent Δ*atpA* mutant: black, darker blue, darker red = wild type; gray, lighter blue, lighter red = Δ*atpA*. a) Specific growth rate of bacterial population calculated by the first derivative of population growth curves. Gray exhibits a larger perturbation from black compared to light blue from dark blue and light red from dark red, indicating the similarity of wild type and Δ*atpA* growth curves during slow growth (in M9CG and M9G). b) Maximum growth rate and saturation point of growth of wild type and Δ*atpA* compared in three different media. The saturation of wild type and mutant display similarities during slow growth. c) Oxygen consumption rate and extracellular acidification rate of wild type and mutant compared in three different media. Δ*atpA* cells consume oxygen and glucose at a higher rate compared to wild type in order to compensate for their lower ATP synthesis rate.

5.4 Deletion of the ATP synthase *atpA* subunit increases death rate in single *E. coli* cells

Differences among the growth curves of bacterial populations may reflect different adjustments in growth on a single-cell level, i.e., a difference in cell mass or cell proliferation rate of *wt* and mutant cells. Alternatively, it could also reflect difference in cell senescence and death rate. To investigate this issue, we examined the growth, proliferation and phenotypic behavior in senescence and death of *wt* and *ΔatpA* mutant *E. coli* cells at the single-cell level when cells were trapped individually in a microfabricated array of channels (i.e., in a "mother machine")[127]. In such experiments, we selected one colony from a plate culture, allowed it to grow in the desired medium, and then loaded the cells into the machine when the population was in its exponential growth phase (for details of experimental procedures see Section 5.1.6). As a result, all cells in each experiment were the descendants of one single mother cell; however, their relationship at the time of the measurements was distanced and unknown to us. After enough cells were trapped, we fed the cells with one of the media used in the previous sections (5.2 and 5.3) and acquired their images in phase contrast and fluorescence modes until they stopped proliferating, and their fluorescence decayed to non-detectable levels. Images taken over several days were analyzed image by image to obtain the data presented here. If a cell was trapped at the end of a channel in the first frame and was still elongating during the last 10 frames, it was considered to be alive at the end of the experiment, otherwise it was considered dead. Sometimes a cell would enter a channel in the middle of the experiment, divide for a while and then die; these cells were also considered in our statistics since.

To investigate whether the death rate of cells can explain the difference in the population growth observed in Figure 68, we used the survival function, which is a measure of how many cells are still proliferating after a certain amount of time. This measure cannot be combined from

different experiments due to the unknown differences of the initial conditions and minor differences in nutrition, so the results of this analysis are presented here for one example experiment in each strain and each medium. The first columns of Figure 69 and Figure 70 show such measurements for wild type and *ΔatpA* cells, respectively. The cells are considered dead after they stop replicating in these graphs. These data points could all be well-fitted to a Gompertz distribution, $e^{-\frac{a}{b}(e^{bt}-1)}$, where a is a constant, b is the death rate and t is time. We also calculated the probability of cell death during a certain time, namely the hazard function. A hazard function that increases with time indicates increased death probability with time, which signifies aging. Although molecular mechanism of aging in bacteria has yet to be identified, multiple previous studies have demonstrated aging in *wt E. coli* by means of the increase in their hazard function[54].

The hazard functions for cells in different media are presented in the right columns of Figure 69 and Figure 70 for wild type and *ΔatpA* cells, respectively. These plots demonstrate aging in our cultures, in agreement with published literature, irrespective of the medium or the strain used. In addition, our results show that the deletion of the *atpA* subunit of the ATP synthase does not alter the behavior of the hazard accumulation function in time. In both, *wt* and *ΔatpA* strains, the hazard accumulation increases linearly in time. However, comparing the death rate of the *ΔatpA* cells to that of the *wt*, represented by the parameter *b* in the Gompertz distribution function, which is obtained by fitting our data to the Gompertz model, reveal a significantly higher death rate for the mutant strain. Figure 71 illustrates a comparison of the values for both strains in all three environmental conditions. It is evident that the death rate of cells is much higher when the *atpA* subunit of the ATP synthase is deleted. Also, the death rate of both strains increases as the nutrients become more limited in the growth medium. A negative correlation between the death

rates and the maximum growth rates found for the population in Figure 68b suggests that a higher death rate in single cells results in a lower maximum growth rate of the population. r

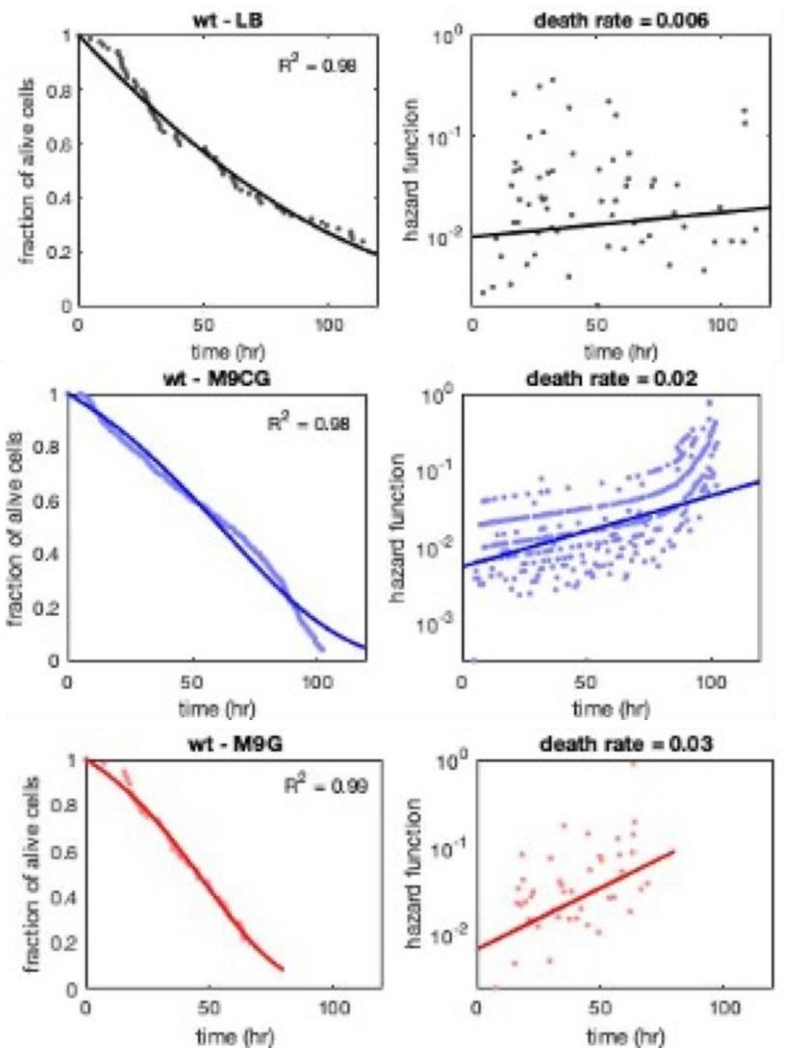


Figure 69 Survival functions and Hazard functions of MG1655 wild type cells in three different media. Only cells that are still proliferating are considered as survived cells. Black represents the medium LB, blue represents M9CG, and red represents M9G. The dots are the actual data, and solid lines are the fits. The survival function has been fitted to a Gompertz distribution with the equation *exp(-(a/b)(exp(bt)-1))*, the value of *b* from this fit has been used to plot the exponential function (solid line) on the hazard function graph with the equation ae^{bt}. The R^2 values are the coefficient of determination of the Gompertz fit. The value of death rate increases as the nutrition is limited.

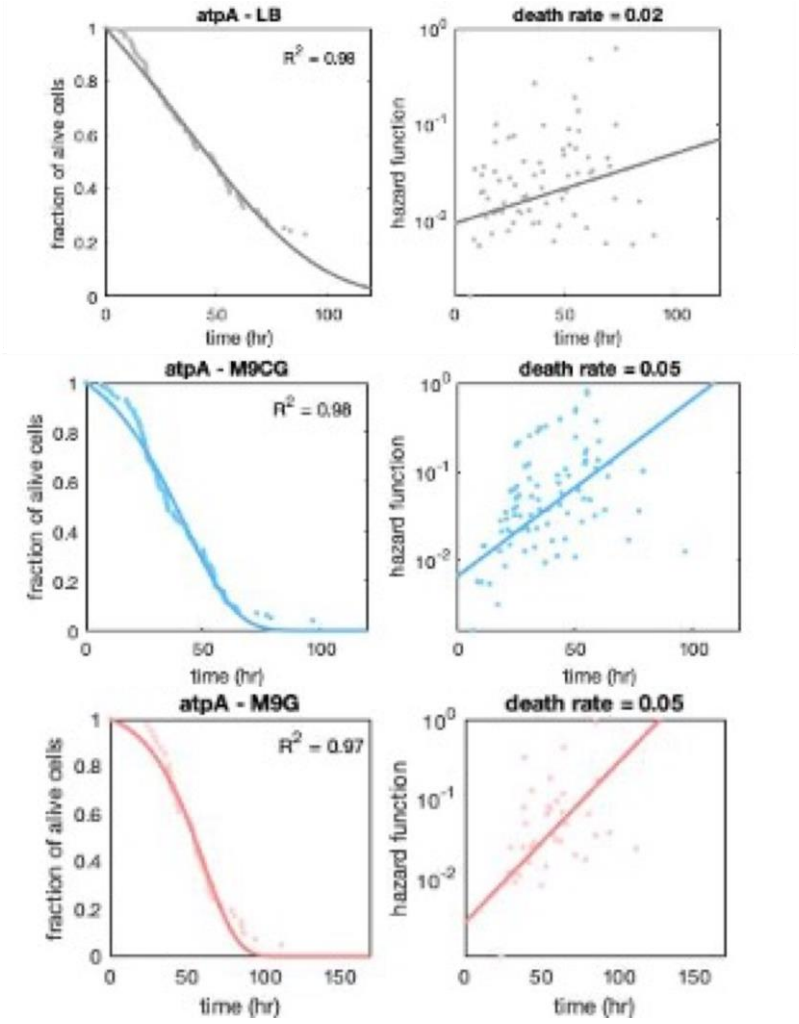


Figure 70 Survival functions and Hazard functions of MG1655 ΔatpA cells in three different media. Only cells that are still proliferating are considered as survived cells. Grey represents the medium LB, light blue represents M9CG, and light red represents M9G. The dots are the actual data, and solid lines are the fits. The survival function has been fitted to a Gompertz distribution with the equation *exp(-(a/b)(exp(bt)-1))*, the value of *b* from this fit has been used to plot the exponential function (solid line) on the hazard function graph with the equation ae^{bt}. The R^2 values are the coefficient of determination of the Gompertz fit. The value of the death rate is higher when the nutrition is more limited.

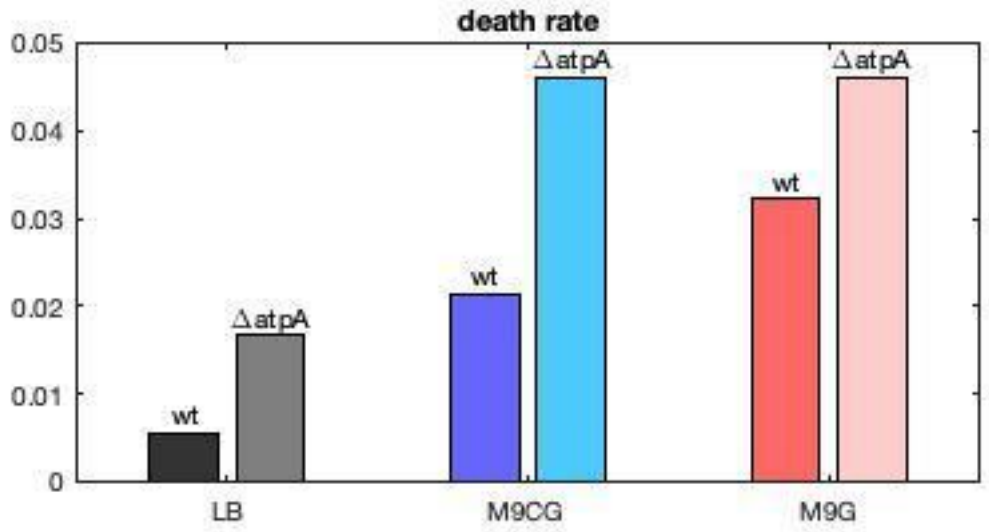


Figure 71 Values of death rate for MG1655 wild type and *ΔatpA* cells. These values were extracted from the Gompertz fits of Figure 69 and Figure 70. It is evident that cells without the *atpA* subunit of the ATP synthase have a higher death rate compared to the wild type. Also, when the nutrition is more limited, the death rate becomes higher. A negative correlation between these values and the population maximum growth rate values represented in Figure 68b is observed.

5.5 Mutant cells exhibit a higher post-replicative lifetime compared to wild type cells

Another major revelation of our single cell analysis was that cells did not disintegrate as soon as they stopped proliferation. To better identify the exact time of death relative to the cell's replicative lifetime, we utilized a commonly used commercial bacterial viability kit (*Bac*Light™, Thermofisher) in our experiments. The red dye propodeum iodide from this kit was added to the medium being pumped into the mother machine at all times. This dye can only enter the cell and stain it when the plasma membrane has been disintegrated. To our surprise, the dye did not stain cells right after they stopped proliferating, but several hours later. We termed this segment of the cell's lifetime as 'post-replicative lifetime' (PRL) and the time where the cell is proliferating as

'replicative lifetime' (RL). The values for these two parameters were measured in all experiments and are represented in Figure 72. Note that the replicative lifetime presented in this figure is not a real parameter, because at the time that the cells enter one of the traps in our experimental setup, the number of replications since its birth is not known. However, due to the high number of data points in each experiment, the distribution of these values should be a good representation of the entire population. In addition, with respect to the conditions of our experiments, comparison between wild type and mutant cells is valid even if the values themselves do not have a significant meaning.

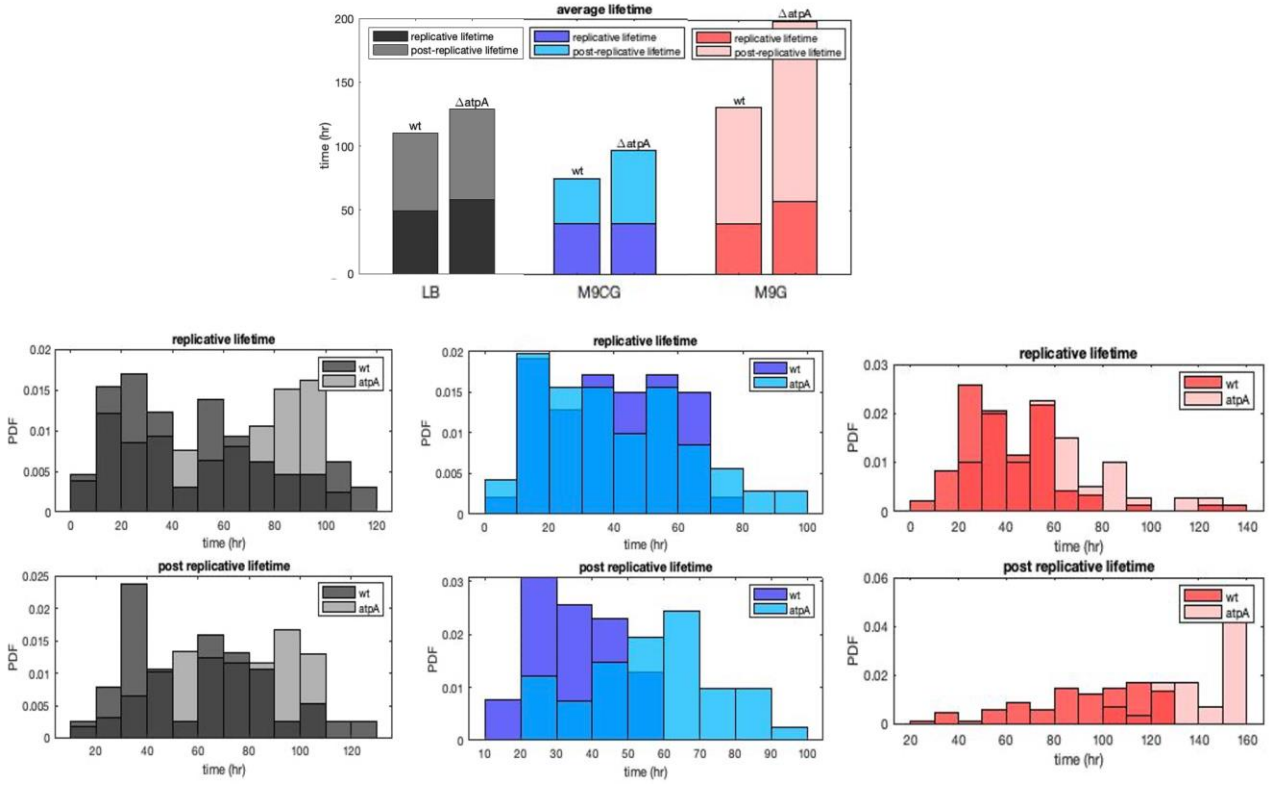


Figure 72 A comparison of replicative lifetime and post-replicative lifetime between MG1655 wild type and *ΔatpA* cells. The two strains show more or less similar distributions of replicative lifetime, however an increase is observed in the post-replicative lifetime of *ΔatpA* cells. Three colors refer to three media: black is LB, blue is M9CG, and red is M9G.

A comparison of the replicative lifetime distributions of the *wt* and the *ΔatpA* strains, reveals that they exhibit almost the same range, albeit with distinct distribution shape. Wild type cells tend to have a higher fraction of cells dying earlier in the experiment, while the distribution of mutant cells is more uniform. Notice that these values are the same used in Figure 69 and Figure 70, which show a higher death rate for mutant cells. It is also worth noting here that the distribution of replicative lifetime for wild type and mutant cells looks very similar in M9CG, where they also exhibit similar population growth (Figure 66).

An examination of the post-replicative lifetime distributions in Figure 72 shows that the mutant cells have significantly higher values in M9CG and M9G. This indicates that while mutant cells do not significantly differ in their replicative lifetime from *wt* cells, they are able to maintain their cellular integrity for a longer time.

5.6 *E. coli* cells exhibit three death phenotypes

A thorough examination of the images acquired from single cells described in the previous section reveal that individual cells have different fates and growth rates depending on their environmental conditions. Here, we focus on the modes of senescence and death that the different cells exhibit. A subset of cells lose one of the plasmids we introduced, and die as a result of losing the antibiotic resistance gene contained in that plasmid, and which allows the cells to survive the antibiotic present in the medium. Since the death of such cells is due to our experimental conditions, we did not consider this mode of death as a phenotype in our analyses. We organize the rest of the cells into three phenotypes based on the death mode they exhibit. In this section we describe the different phenotypes of cell death identified in our experiments.

Our results show that *E. coli* cells display three distinct death phenotypes. Figure 73 presents an example of each of these phenotypes. In phenotype I- cells lose their chromosome during division, however, the cell can still perform some activities that are not reliable on the chromosome (Figure 73a). This activity in bacteria with no chromosome has also been reported previously[128]. In phenotype II- cells preserve their chromosome throughout their entire life, but they enter a non-dividing phase before lysis (Figure 73b). In phenotype III – cells behave the same as in phenotype II, with the difference that at the time of their last division, they become filamented. We consider any cell whose length exceeds twice the average size in the same strain and medium to be filamented (Figure 73c). Most cells in all experiments filamented at some point during their replicative lifetime, however phenotype III only includes cells that remained long after their last division. Also, none of type I cells were filamented during their last division. In instances where the cells became much longer than the channel length (30 μm), they would escape from the channel, so our statistics do not include that subset of the population. In Figure 73 the time before the dashed line represents the replicative lifetime of the cell, and the time between the black and the red dashed lines represents the post-replicative lifetime.

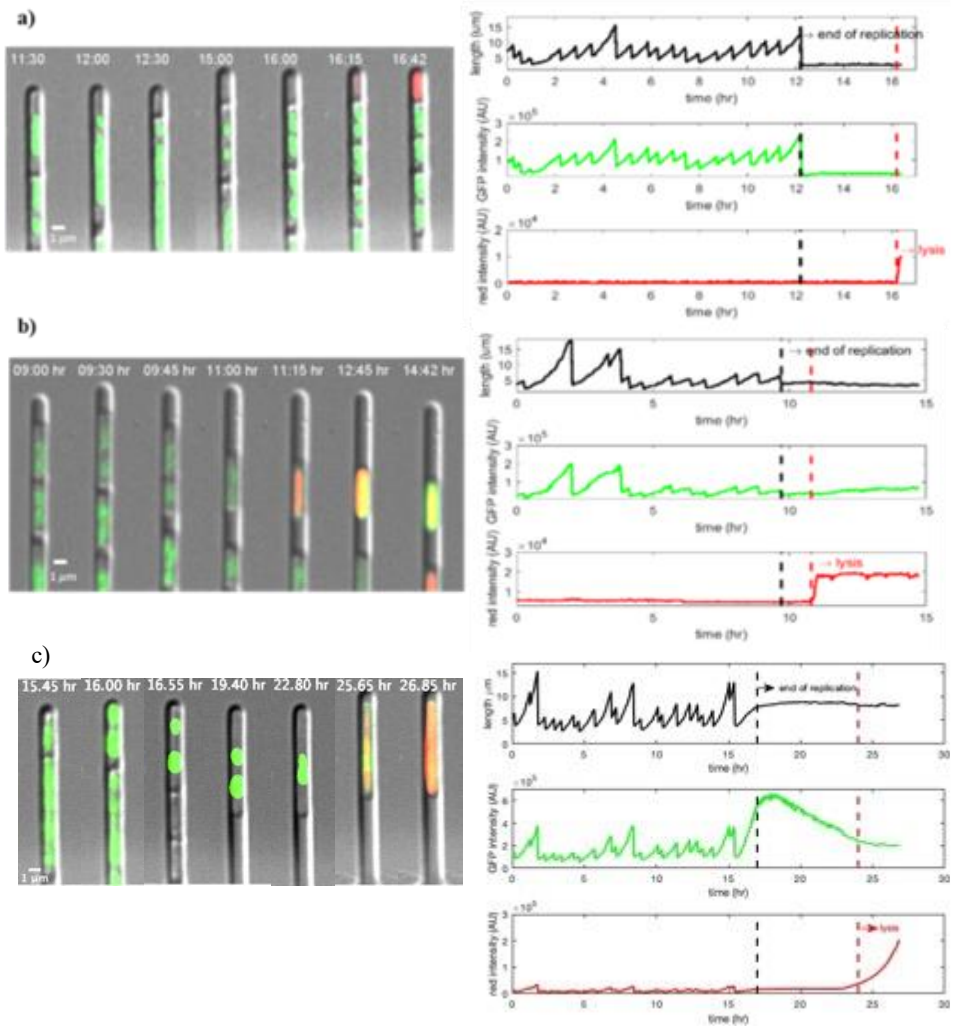


Figure 73 Growth and death phenotypes of individual *E. coli* cells. Superimposed images of differential interference contrast (DIC), green, and red fluorescence. Cells express GFP-Fis upon induction with IPTG which bind to the chromosome making it visible, hence the green color in these figures represents the cell nucleoid[129]. Propidium iodide is present in the medium, which stains lysed cells red. All three example cells are wild type BW25113 in LB. (a) An example of a Type I cell. The last division of this cell's lifetime occurs at 12:30 in which a part with no chromosome is separated. This cell can hold on to its structure for more than three hours, when at 16:15 the cell membrane is compromised and the dye can enter (lysis). (b) An example of a Type II cell. This cell divides at 09:45 for the last time. It still contains a chromosome but no change in length is visible after this division. Lysis occurs at 11:15. (c) An example of Type III cell. The last time-point that the cell is changing in length is at 16:55 where there are two separate chromosomes in the cell. The cell then holds its structure until time 24:00 where the red die can enter the cell.

5.7 Single cells switch their phenotype to adapt to new conditions

The fraction of each phenotype in the two strains growing in different media are presented in Table 5 (for details see Appendix Table 2). Our results indicate that the fraction of phenotype II within the population is increased upon deletion of the α-subunit of ATP synthase. Similarly, *wt* cells switch to phenotypes II and III in an environment with lower nutrients (M9CG or M9G). Thus, we can conclude that the transformation from phenotype I to phenotypes II and III is the source, or one of the sources, of the differences observed on a population level and is a direct result of bacterial adaptation to the environment. In addition, our observations revealed that some cells would filament right before their growth stops (phenotype III), while others would not (phenotypes I and II). It is known that filamentation is an SOS response of the cell, but the details of how and when a cell enters this phase is not completely known[56].

Table 5 Percentage of each death phenotype present in two strains and in three growth media. Wild type *E. coli* cells revealed three different death modes, while Δ*atpA* cells revealed only two.

	Wild type			Δ*atpA*		
	Type I %	Type II %	Type III %	Type I %	Type II %	Type III %
LB	20.4	59.2	20.4	0	76.3	23.7
M9CG	16.3	65.5	18.2	0	83.6	16.4
M9G	10.7	17.7	71.6	0	56.4	43.6

We next questioned whether any of the discovered phenotypes expresses spontaneous death. The cells that were still alive at the end of the experiment were removed from the statistics, and we divided the rest into three categories and calculated the survival function for each phenotype. Figure 74 displays the hazard function for each phenotype in three different media for wild type cells. It can be seen that all three types exhibit aging, but with different death rates. The

death rate of phenotype I seems to be the highest, which means that cells that lose their DNA tend to die much faster than cells that do not lose their DNA. Figure 75 displays the same data for *ΔatpA* cells. In most cases, type III cells have a higher death rate compared to type II cells. Figure 76 represents all death rates of both strains in three media. A comparison of these values shows that as mentioned before, type I cells tend to have a higher growth rate compared to type II and type III. This figure also reveals that wild type and *ΔatpA* cells exhibit different behaviors in different media. For example, when *ΔatpA* cells are in M9G medium (low nutrition), type II exhibits a lower death rate compared to type III, but for wild type cells in the same medium, type II cells exhibit almost the same rate compared to type III. These differences could account for the differences observed in the population growth seen in Figure 66.

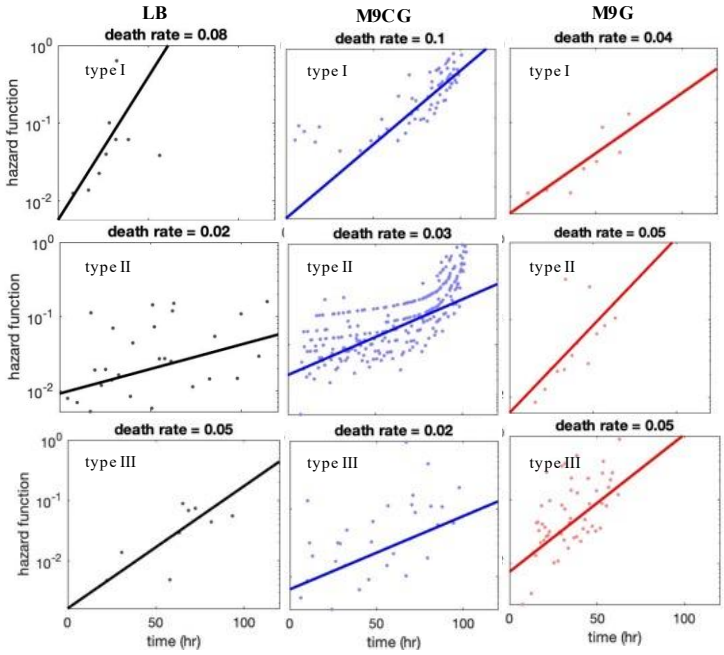


Figure 74 The hazard functions for wild type cells in three different media. Each phenotype was considered as a separate population which their survival function goes from 1 (all cells alive) to 0 (no cells alive), then the hazard function was calculated from it as described in Section 5.4. Colors correspond to different media, black is LB, blue is M9CG, and red is M9G. All three phenotypes in all three media exhibit signs of aging but with different death rates.

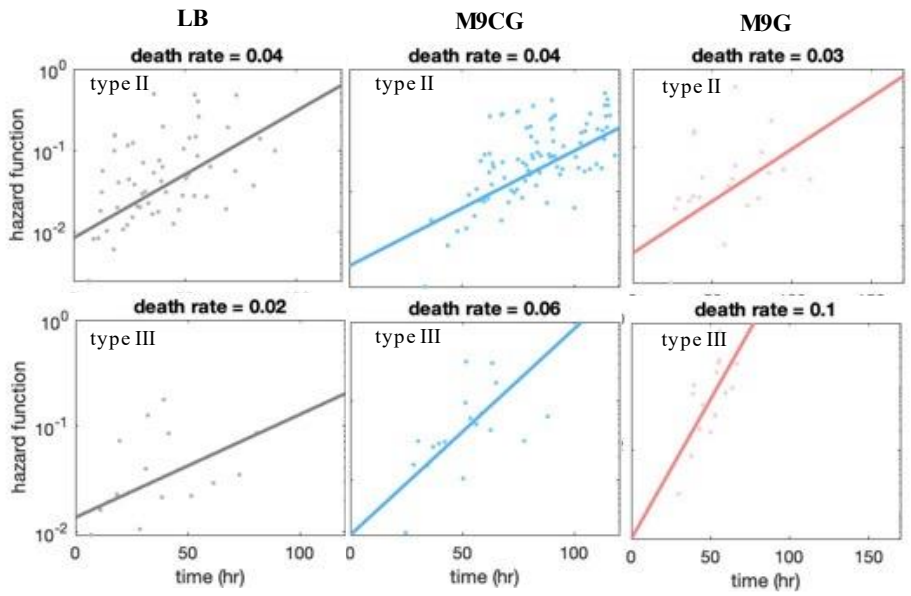


Figure 75 The hazard functions for *ΔatpA* cells in three different media plotted as described in Figure 74. Mutant cells only exhibit two phenotypes. Colors represend media, black is LB, blue is M9CG, and red is M9G.

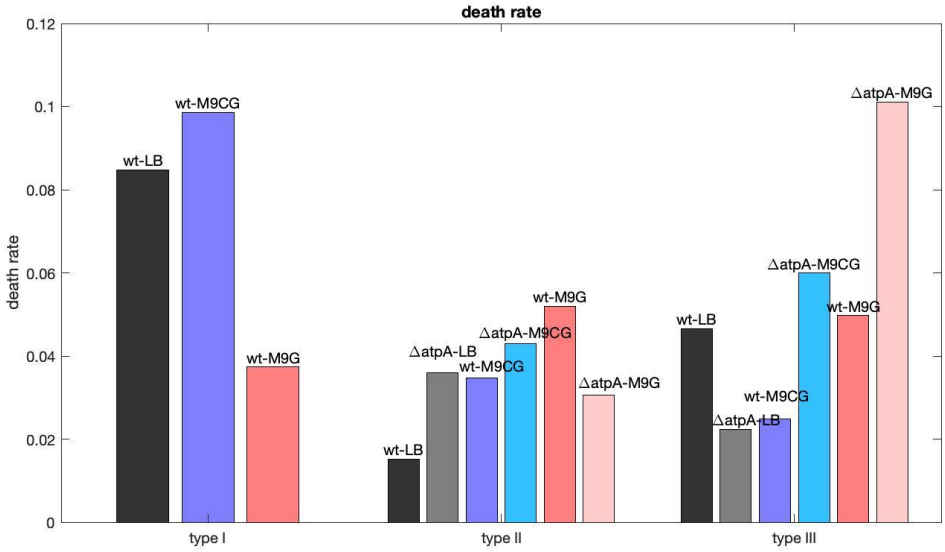


Figure 76 The measured death rate for each phenotype in each strain in each medium. Type I cells exhibit a higher death rate compared to the other two phenotypes. Wild type and *ΔatpA* cells behave differently in different media. Colors represend media, black is LB, blue is M9CG, and red is M9G. Darker colors are wild type and lighter colors are *ΔatpA* cells.

The distributions of replicative and post-replicative lifetimes of each phenotype were also determined, and they are illustrated in Figure 77.

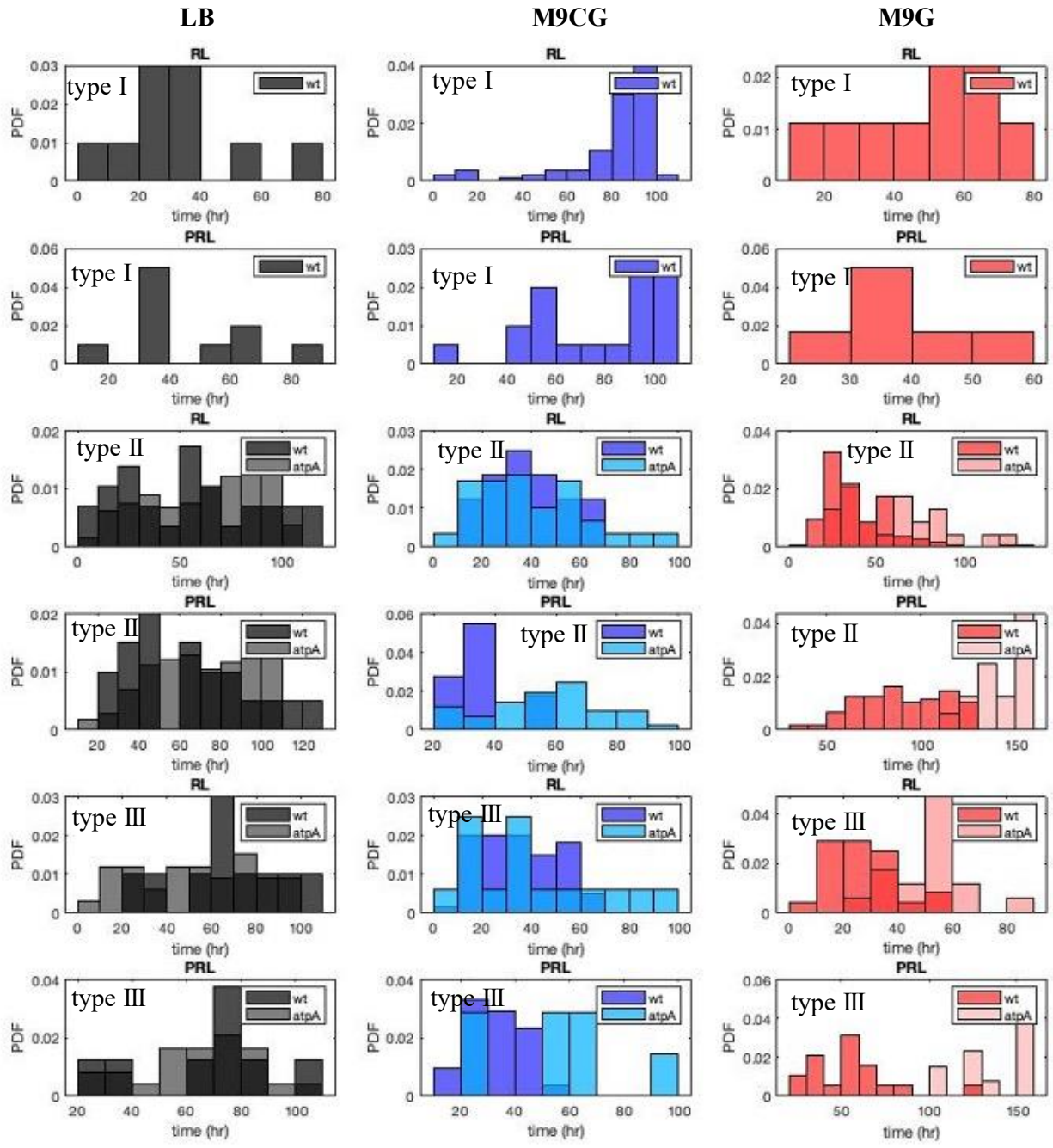


Figure 77 Replicative lifetime (RL) and post-replicative lifetime (PRL) of each phenotype in each strain in each medium. The distributions for wild type and mutant cells look similar in LB, but in M9CG and M9G

mutant cells exhibit a higher percentage of the population with long lifetime. Colors represend media, black is LB, blue is M9CG, and red is M9G. Darker colors are wild type and lighter colors are *ΔatpA* cells.

These distributions seem to behave similarly for wild type and *ΔatpA* cells in LB (rich medium), but in M9CG and M9G the mutant cells exhibit a higher number of cells with longer lifetimes. Figure 78 compares the averages of these distributions. It can be seen that the average replicative lifetime of *ΔatpA* cells is always lower than wild type cells, except for when they are growing in M9G. Note that all these lifetimes are measured in hours, and it is possible that the results would be different if the lifetime was measured as number of divisions. It is interesting to plot these graphs in generation instead of time, which will be accomplished in a future study by our group.

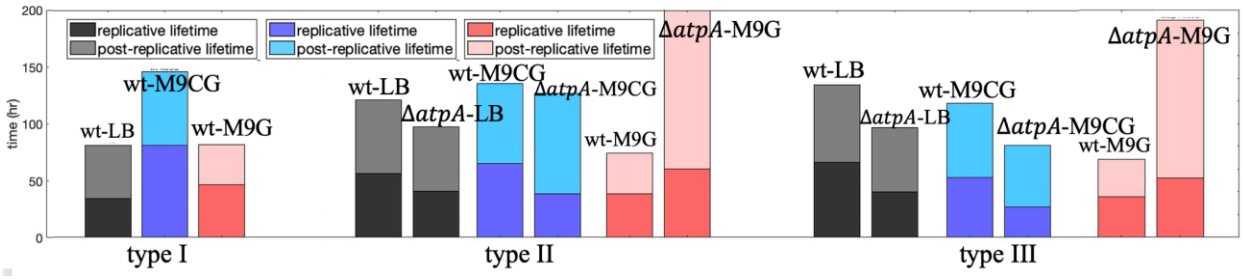


Figure 78 The average replicative and post-replicative lifetimes of each phenotype in each strain in each medium. These values are the averages of the distributions in Figure 77. Mutant cells exhibit shorter average lifetimes compared to wild type cells, except for in M9G. Colors represent media, black is LB, blue is M9CG, and red is M9G. Darker colors are replicative lifetime and lighter colors are post-replicative lifetime.

5.8 Conclusions

The main goal of this chapter was to link the metabolic efficiency of cells to their growth rate and death. For this purpose, we performed two types of experiments, one set of population experiments where a very low concentration of cells were grown in a medium until saturation, and

one set of single-cell experiments where the experimental conditions were favorable and allowed a cell to grow forever, however, we observed that this does not happen and that all cells eventually die.

Moreover, we find that *ΔatpA* cells in all media display a higher death rate compared to *wt*, which can be due to their lower metabolic efficiency. Our results show that death rates were always higher, and maximum growth rates were always lower, in strains and media that revealed lower OCR and EACR, which again points to the fact that the cells with lower metabolic efficiency die sooner.

6.0 Discussions and future directions

This research provides a better understanding of the cellular growth and its control affect size homeostasis and senescence in bacteria organisms, which are fundamental processes essential for all life forms, and central problems in biophysics research. In the current study, we utilize recent technical developments that have made it possible to obtain large number of precise measurements of cell growth and gene expression dynamics at the single-cell level, in order to develop a better quantitative characterization of bacterial growth and division dynamics, and lysis processes.

In chapter 2.0, we discussed one of these measurement methods applied in this research to obtain the main experimental data, which is based on the use of a microfluidic device known as the "mother machine"[40]. Using this experimental setup, we were able to obtain high-throughput, high-quality measurements of cell growth and division dynamics. Analyzing such large quantities of data in a meaningful manner to infer phenomenological models and molecular mechanisms that govern cellular behavior remains a challenging task. Much of the effort is usually guided by preconceived models of cell growth and presumed roles of various factors, which are then tested in the experimental data[13,118]. Such an approach could limit the ability to uncover hidden information as it narrows the scope of the examination of the data. In Chapter 3.0, we have employed a new method in analyzing the dynamics of single bacterial cells, which aims to uncover dependencies among different measurable quantities without any prior assumptions about underlying mechanisms. Our method, which we termed "constrained multivariate regression", is similar to both standard "multiple regression" and "multivariate regression". Our analysis reproduces dependencies that were reported in previous publications, including those found in

phenomenological models such as the adder model[10,12,130]. Our findings emphasize a more general aspect of this approach. A search for mechanisms controlling size regulation of dividing cells is only one example of recent quantitative studies of complex biological processes. Looking for *the* underlying mechanism of a biological phenomenon, such as the bacterial growth control, can lead to proposing preconceived models about the nature of the mechanism. In reality however, many mechanisms can contribute to the same phenomenon, and their relative contributions can also change depending on various conditions. It is thus important, not to be restrained by such preconceived models and to use more "agnostic" approaches for quantitative analyses, such as the one presented in this chapter.

This method provided us with the same qualitative theories from several repeat experiments. However, the values that we achieved varied quantitatively among experiments. These variations could be due to cryptic difference between cell lineages and their history, such as growth time and final cell concentration reached prior to loading the cells into the traps. They could also result from experimental conditions beyond our control. The latter illustrates the presence of possible systematic errors even in well-controlled experiments. This may include small variation in the width of the microchannels obtained from the microfabrication process[65], the conditions of the media used, and the temperature dynamics in the incubator. Another potential source of variation is the build-up of microenvironments in those microchannels over time. As often is the case, answering quantitative questions more precisely calls for improving the experimental approach. To remove possible microenvironments, more sophisticated microfluidic devices can be used, such as replacing one side of the microchannels with a semipermeable membrane through which nutrients can be exchanged more efficiently[131] as shown in Figure 79.

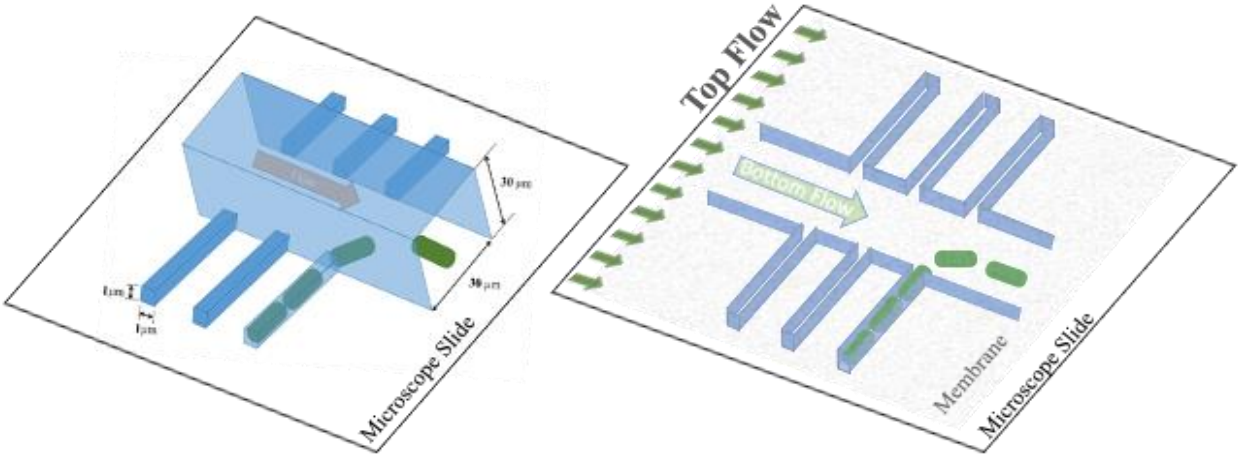


Figure 79 The left device depicts a current model of the mother machine where cells are fed through a wide channel. The right device is a possible model for a new device, where a semipermeable membrane is placed on the entire device so that the cells can be fed through a flow that moves above the channels, and the offsprings can be washed out through a flow that is under the membrane.

Another important result of our analysis in Chapter 3.0 is that two sister cells born from the same mother cell exhibit distinct growth rates; the cell that receives a smaller fraction has a higher growth rate. The main theory of symmetrical division of bacteria states that when a cell divides, each daughter cell receives half of the molecular content of the mother cell. Since the molecular content of the cell is distributed uniformly in the cell volume, each daughter receives the same concentration of molecules with the exemption of DNA. Multiple chromosome replication forks are initiated in each cell while growing in length, and once two are complete, the cell divides and allocates one complete copy of the chromosome, with all the ongoing replication forks attached to it, to each of its daughters. The number of replication forks that each daughter receives is the same, which results in higher DNA concentration in the smaller sister cell. This in turn would lead to a faster production of proteins and other cellular building blocks and thus the

smaller sister grows faster than its larger sister. In Chapter 4.0, we tested this hypothesis by forcing the cell to grow without division with and without DNA replication. This increased the span of DNA concentration in the cells and provided a wider dynamical range of cellular growth. Examining the relationship between DNA concentration and growth rate using this method revealed a clear linear correlation between the two, suggesting that indeed, higher DNA concentration leads to faster cellular growth.

To uncover more details of the molecular mechanism of cellular growth the effect of other cellular components involved in producing building blocks for the cell, on the growth rate should be investigated. The effect of ribosomes copy number on the cell growth rate has been speculated for a long time, however, it has never been measured at the single cell level. In addition, since the production of all important ingredients in the cell starts from the action of the RNA polymerase (RNAP) on the DNA, a higher concentration of DNA to RNAP could lead to a faster cellular growth and so a more detailed investigation on this matter could reveal interesting results. Such investigations combined with the results of our study here, could lead to the establishment of a quantitative mathematical link between the cellular growth regulation mechanism and the observed phenomenological feedback depicting cell size homeostasis, which is currently lacking.

In addition to cellular growth, cellular death is also one of nature's fundamental realities that is still subject to intense research efforts and the question of why all organisms undergo senescence and death is still challenging. Previous studies have assigned various reasons to cellular aging such as accumulation of damage[132], genome instability[133], and telomere maintenance in eukaryotes[134]. In Chapter 5.0, we explored how metabolic efficiency can contribute to cellular aging and death. We demonstrated that the model bacterium, *E. coli*, displays three alternative aging phenotypes, and that the majority of cells enter into a state that is phenotypically similar to

eukaryotic cells' replicative senescence before lysis. We were also able to reveal signs of aging in wild type cells as well as a mutant where a subunit of the ATP synthase was deleted and show that in the absence of sufficient nutrients or in the case of a mutation, cells can adapt to their new environment by changing the ratio of phenotypes in the population. In conclusion, lowering the metabolic efficiency of cells by deleting a subunit of one of its most important enzymes (ATP synthase) does not lead directly to cellular death, however, it does affect cellular aging. Even though *E. coli* is a prokaryotic organism, where aging is different, using the toolkit of systems biology, we can extend our understanding of how cells' metabolic network organization and function influence their aging phenotype in bacteria to be applied to other eukaryotic cells as well as human aging-related diseases such as Alzheimer's and heart diseases.

The ultimate goal of such research studies is to be able to model senescence of *E. coli* and the effects of metabolic efficiency on aging, and to be able to obtain concrete conclusions on how the processes of aging and death are occurring. This model can then be used for a regression analysis, which will examine the correlation between the obtained parameters and their corresponding noises on the last generations before entering non-replicative lifespan. This will provide insight into whether the cell has a memory of how long it has lived, and as a consequence is determined to age and die. For a better understanding on the molecular mechanisms directly affecting cellular death, the content of a cell can be examined to observe whether an accumulation of damage appears, and in the case of its presence the aggregation could be identified by using dyes with the ability to tag proteins and DNAs, to recognize its content. In addition, the diffusion of the aggregation and origin of accumulation can be studied using single molecule tracking. The effect of single cell growth parameters (such as growth rate or division time) on the size and location of aggregations is also an interesting subject to investigate. Using a different microfluid

device where a cell can be trapped from its birth to death could also provide more insight into the process of aging. Genome-scale models such as flux balance analysis (FBA) are well-established tools for studying the metabolic state of cells. This modeling approach might be able to identify metabolic enzymes whose deletion or inactivation affect the replicative and non-replicative lifespans of *E. coli* cells, but do not kill them, such as the one studied here.

Appendix A Fitting growth curves

The fitting parameters are: the initial size of each cell cycle, $x_1^A, \cdots, x_N^A$ and $x_1^B, \cdots, x_N^B$; the growth rate of each cell cycle, $\alpha_1^A, \cdots, \alpha_N^A$ and $\alpha_1^B, \cdots, \alpha_N^B$; the time of division at the end of each cell cycle, $t_1^A, \cdots, t_N^A$ and $t_1^B, \cdots, t_N^B$; the beginning time of the first cell cycle, t_0^A; and finally, an overall length offset, z. Thus, the duration of each cell cycle is given by $\tau_n^A = t_n^A - t_{n-1}^A$ for cell A and $\tau_n^B = t_n^B - t_{n-1}^A$ for cell B. Among the fitting parameters, $t_1^B, \cdots, t_N^B$ as well as t_N^A and t_0^A (the latter is degenerate with x_1^A) are not affected by the constraint, so we set those to be the mid-point of the two consecutively measured time points between which a division happened. The rest $5N$ parameters are fitted using a least-squares cost function,

$$L = \sum_{a=A,B} \sum_{n=1}^{N} \sum_{i=1}^{l_n^a} \frac{1}{2} \left(X_{n,i}^a - x_n^a \, \mathrm{e}^{\alpha_n^a \left(t_{n,i}^a - t_{n-1}^a \right)} - z \right)^2$$

with constraint equations,

$$x_n^A + x_n^B = x_{n-1}^A \, \mathrm{e}^{\alpha_{n-1}^A \left(t_{n-1}^A - t_{n-2}^A \right)} \quad \text{for } n = 2, \cdots, N.$$

The cost function is minimized numerically using the sequential quadratic programming method.

An example of the fitted cell growth curves is shown in Appendix Figure 3. Note that the consecutive cell cycles of cell A are adjacent in time, whereas those of cell B can overlap or have gaps. From the fitted parameters, we could calculate the other cell cycle variables described in the main text. In particular, $\chi_n^a = \log x_n^a$ and $\phi_n^a = \alpha_n^a \, (t_n^a - t_{n-1}^A)$. Besides, the minimum value of the cost function represents the variance of the residual from the curve fitting, which gives an estimate of the measurement error:

$$\sigma^2 = 2L_{\min} \Big/ \sum_{a=A,B} \sum_{n=1}^{N} l_n^a$$

Algorithm written in Python 3.5[135] by BingKan Xue:

```
# read raw data on cell length vs time: A (mother) and B (daughter)
# locate division time and label cell cycle
# extract growth factor, division fraction, and cycle duration

import sys, os, glob
import numpy as np
import scipy.optimize as opt
import matplotlib.pyplot as plt

input_dir = '../SalmanLab/Traces_2cells'
input_name = ''
output_name = 'trace(crct)'
output_name1 = 'extract(crct)'
output_dir = 'results6_042817'
if not os.path.isdir(output_dir):
    os.mkdir(output_dir)
file_list = glob.glob('%s/%s*.txt' % (input_dir, input_name))
skip = 1    # number of rows to skip in the data file
num_all = []

for f in sorted(file_list):
    # read file
    filename = os.path.basename(f)[len(input_name):-4]
    dat = np.genfromtxt(f, delimiter='\t', skip_header=skip)
    print '\nfile: %s' % filename

    # extract time and size for both cells
    crop = ~np.isnan(dat[:,0])
    timeA = dat[crop,0]
    sizeA = dat[crop,1]
    if np.isnan(sizeA).any():
        raise RuntimeError('NaN found in data')
    crop = ~np.isnan(dat[:,4])
    timeB = dat[crop,4]
    sizeB = dat[crop,5]
    if np.isnan(sizeB).any():
        raise RuntimeError('NaN found in data')

    # locate divisions
    divA = np.diff(sizeA) < -1    # size decrease
    begA = np.nonzero(divA)[0][:-1] + 1    # beginning of each FULL cycle
    endA = np.nonzero(divA)[0][1:]    # end of each FULL cycle
    num = len(begA)    # number of cell cycles recorded
    num_all.append(num)
    print 'number of cycles: %d' % (num)
    sizeB[np.isnan(sizeB)] = 99.    # cell B data may contain NaN (cell escapes)
    divB1 = np.diff(sizeB) < -0.6    # size decrease
    divB2 = np.diff(timeB) <= 0      # overlapping cycles
    divB3 = np.diff(timeB) >= 0.3   # gap between cycles
    divB = np.any(np.vstack([divB1, divB2, divB3]), axis=0)
    begB = np.nonzero(divB)[0][:-1] + 1    # beginning of each FULL cycle
    endB = np.nonzero(divB)[0][1:]    # end of each FULL cycle
```

```
53.
54.     # fit exponential growth for all cycles
55.     start = np.zeros(num)     # start time of each cycle
56.     slopeA = np.zeros(num)    # slope of each cycle
57.     duratA = np.zeros(num)    # duration of each cycle
58.     initA = np.zeros(num)     # initial size of each cycle
59.     finaA = np.zeros(num)     # final size of each cycle
60.     growA = np.zeros(num)     # growth factor during each cycle
61.     fracA = np.zeros(num)     # division fraction (linear) BEFORE each cycle
62.     slopeB = np.zeros(num)    # slope of each cycle
63.     duratB = np.zeros(num)    # duration of each cycle
64.     initB = np.zeros(num)     # initial size of each cycle
65.     finaB = np.zeros(num)     # final size of each cycle
66.     growB = np.zeros(num)     # growth factor during each cycle
67.     fracB = np.zeros(num)     # division fraction (linear) BEFORE each cycle
68.     both = np.zeros(num)      # sum of initial sizes BEFORE each cycle
69.
70.     # initial guess by fitting exponential curves independently for each cycle
71.     for n in xrange(num):
72.         # cell A
73.         bnA = begA[n]
74.         enA = endA[n]
75.         indA = np.arange(bnA, enA+1)
76.         start[n] = 0.5 * timeA[bnA] + 0.5 * timeA[bnA-1]  # extrapolate
77.         if len(indA) < 3:    # too few data points
78.             print 'too few points, skip fitting A'
79.         else:
80.             # fit exponential curve
81.             def fun(t, a, b):
82.                 return a * np.exp(b * (t - start[n]))
83.             fit, cov = opt.curve_fit(fun, timeA[indA], sizeA[indA], p0=(3., 1.))
84.             initA[n], slopeA[n] = fit
85.         # cell B
86.         bnB = begB[n]
87.         enB = endB[n]
88.         indB = np.arange(bnB, enB+1)
89.         if sizeB[enB] == 99.:    # incomplete cycle (cell escaped)
90.             print 'incomplete cycle, skip fitting B'
91.         elif len(indB) < 3:    # too few data points
92.             print 'too few points, skip fitting B'
93.         else:
94.             # fit exponential curve
95.             def fun(t, a, b):
96.                 return a * np.exp(b * (t - start[n]))
97.             fit, cov = opt.curve_fit(fun, timeB[indB], sizeB[indB], p0=(3., 1.))
98.             initB[n], slopeB[n] = fit
99.     missB = np.nonzero(initB == 0.)[0]    # skipped cycles
100.
101.     # optimize variables under length constraint
102.     def growth(time, init, slope, start, offset):
103.         """
104.         exponential growth function.
105.         Input:
106.         time: real, or 1-d array, time points to evaluate size
107.         Parameters:
108.         init: real, initial size at birth
109.         slope: real, exponential growth rate
110.         start: real, start time of cell cycle
111.         offset: real, due to measurement error
112.         Return:
113.         size: real, or 1-d array, estimated size at given time points
```

```
114.         """
115.         size = init * np.exp(slope * (time - start)) + offset
116.         return size
117.
118.     def loss(initA, initB, slopeA, slopeB, start, offset):
119.         """
120.         loss function for least-square fit.
121.         Input:
122.         initA/B: 1-d array, initial size of cell A/B at start of each cycle
123.         slopeA/B: 1-d array, growth rate of cell A/B during each cycle
124.         start: 1-d array, start time of each cycle for both cells A&B
125.         offset: real, measurement error
126.         Return:
127.         sqls: real, square loss between data and estimated sizes
128.         """
129.         sqls = 0.
130.         pts = 0    # total number of data points
131.         for n in xrange(num):
132.             # cell A
133.             bnA = begA[n]
134.             enA = endA[n]
135.             indA = np.arange(bnA, enA+1)
136.             tA = timeA[indA]
137.             dA = sizeA[indA]
138.             sA = growth(tA, initA[n], slopeA[n], start[n], offset)
139.             sqls += 0.5 * np.sum((dA - sA)**2)
140.             pts += len(indA)
141.             # cell B
142.             bnB = begB[n]
143.             enB = endB[n]
144.             indB = np.arange(bnB, enB+1)
145.             if sizeB[enB] == 99.:    # incomplete cycle (cell escaped)
146.                 continue
147.             elif len(indB) < 3:    # too few data points
148.                 continue
149.             tB = timeB[indB]
150.             dB = sizeB[indB]
151.             sB = growth(tB, initB[n], slopeB[n], start[n], offset)
152.             sqls += 0.5 * np.sum((dB - sB)**2)
153.             pts += len(indB)
154.         if np.isnan(sqls):
155.             raise RuntimeError('loss function is NaN')
156.       # save data
157.     fout = open('%s/%s_%s.dat' % (output_dir, output_name, filename), 'w')
158.     fout.write('# exponential fit of cell length vs time\n' +\
159.                '# overall length offset = %.6f\n' % offset +\
160.                '# standard deviation of residuals = %.6f\n' % resid)
161.     info = 'slope, duration, initial size, final size, ' +\
162.            'growth factor, division fraction:'
163.     info += '\ncell A (mother):'
164.     results = np.vstack((slopeA, duratA, initA, finaA, growA, fracA)).T
165.     np.savetxt(fout, results, fmt = '%.8f', delimiter='\t', header=info)
166.     info = 'cell B (daughter):'
167.     results = np.vstack((slopeB, duratB, initB, finaB, growB, fracB)).T
168.     np.savetxt(fout, results, fmt = '%.8f', delimiter='\t', header=info)
169.     fout.close()
```

Appendix B Constrained multivariate regression coefficients

The algorithm for calculating the coefficients of our constrained multivariate regression model was written in Python 3.5 by BingKan Xue and in MATLAB by Maryam Kohram. The input for these algorithms are the outputs of the code provided in Appendix A.

```
%% 04/12/2018 Maryam Kohram - This program calculates R matrices
clear all
source_dir = '/Users/maryam/Box/Computer
Backup/Research/MotherMachine/regression_analysis/data/TLM_output/text';
source_files = dir(fullfile(source_dir,'*.csv'));
figure(1)
for inp = 1:length(source_files)
    filename = [source_dir filesep source_files(inp).name];
    data{inp} = csvread(filename);

    %% input three variables
    X{inp}(i,1) = data{inp}(i,1);
    Alpha{inp}(i,1)= data{inp}(i,2);
    Phi{inp}(i,1) = data{inp}(i,3);
    %% define two extra variables
    Eta = X{inp} + Phi{inp};
    Psi(i,1) = X{inp}(i+1) - X{inp}(i) - Phi{inp}(i);

    %% examine the temporal correlation between the variables
    %% a) <X_(n+l)X_n>
    m = length(X{inp});
    ave = mean(X{inp});
    v=1;
    lag = 9;
    lag1(:,1) = linspace(0,9,10);
    for s=0:lag
        mult=0;
        summ=0;
        for t=1:m-s
            mult=((X{inp}(t+s)-ave))*((X{inp}(t)-ave));
            summ=summ+mult;
        end
        acf{inp,1}(v,1)=summ/(m-s);
        v = v + 1;
    end
    subplot(3,3,1)
    %plot(acf{inp,1})
    axis([-0.5 10 -0.03 0.035])
    title('$<X_{n+l}X_n>$','Interpreter','latex')
    xlabel('time delay l')
    ylabel('$C_{00}(l)$','Interpreter','latex')
    hold on
```

```
%% b)    <X_(n+l)alpha_n>
m2 = length(Alpha{inp});
ave2 = mean(Alpha{inp});
v=1;
stdalpha = std(Alpha{inp});
stdx = std(X{inp});
for s=0:lag
    mult=0;
    summ=0;
    for t=1:m-s
        mult=((X{inp}(t+s)-ave))*((Alpha{inp}(t)-ave2));
        summ=summ+mult;
    end
    acf{inp,2}(v,1)=summ/(m-s);
    v = v + 1;
end
subplot(3,3,2)
%plot(acf{inp,2})
axis([-0.5 10 -0.008 0.006])
title('$<X_{n+l}\alpha_n>$','Interpreter','latex')
xlabel('time delay l')
ylabel('$C_{01}(l)$','Interpreter','latex')
hold on
%% c) <X_(n+l)phi_n>
m3 = length(Phi{inp});
ave3 = mean(Phi{inp});
v=1;
for s=0:lag
    mult=0;
    summ=0;
    for t=1:m-s
        mult=((X{inp}(t+s)-ave))*((Phi{inp}(t)-ave3));
        summ=summ+mult;
    end
    acf{inp,3}(v,1)=summ/(m-s);
    v = v + 1;
end
subplot(3,3,3)
%plot(acf{inp,3})
axis([-0.5 10 -0.02 0.02])
title('$<X_{n+l}\phi_n>$','Interpreter','latex')
xlabel('time delay l')
ylabel('$C_{02}(l)$','Interpreter','latex')
hold on
%% d) <alpha_(n+l)X_n>
v=1;
for s=0:lag
    mult=0;
    summ=0;
    for t=1:m-s
        mult=((Alpha{inp}(t+s)-ave2))*((X{inp}(t)-ave));
        summ=summ+mult;
    end
    acf{inp,4}(v,1)=summ/(m-s);
    v = v + 1;
end
subplot(3,3,4)
```

```
%plot(acf{inp,4})
axis([-0.5 10 -0.008 0.005])
title('$<\alpha_{n+l}X_n>$','Interpreter','latex')
xlabel('time delay l')
ylabel('$C_{10}(l)$','Interpreter','latex')
hold on
%% e) <alpha_(n+l)alpha_n>
v=1;
for s=0:lag
    mult=0;
    summ=0;
    for t=1:m-s
        mult=((Alpha{inp}(t+s)-ave2))*((Alpha{inp}(t)-ave2));
        summ=summ+mult;
    end
    acf{inp,5}(v,1)=summ/(m-s);
    v = v + 1;
end
subplot(3,3,5)
%plot(acf{inp,5})
axis([-0.5 10 -0.03 0.04])
title('$<\alpha_{n+l}\alpha_n>$','Interpreter','latex')
xlabel('time delay l')
ylabel('$C_{11}(l)$','Interpreter','latex')
hold on
%% f) <alpha_(n+l)phi_n>
v=1;
for s=0:lag
    mult=0;
    summ=0;
    for t=1:m-s
        mult=((Alpha{inp}(t+s)-ave2))*((Phi{inp}(t)-ave3));
        summ=summ+mult;
    end
    acf{inp,6}(v,1)=summ/(m-s);
    v = v + 1;
end
subplot(3,3,6)
%plot(acf{inp,6})
axis([-0.5 10 -0.01 0.013])
title('$<\alpha_{n+l}\phi_n>$','Interpreter','latex')
xlabel('time delay l')
ylabel('$C_{12}(l)$','Interpreter','latex')
hold on
%% g) <phi_(n+l)X_n>
v=1;
for s=0:lag
    mult=0;
    summ=0;
    for t=1:m-s
        mult=((Phi{inp}(t+s)-ave3))*((X{inp}(t)-ave));
        summ=summ+mult;
    end
    acf{inp,7}(v,1)=summ/(m-s);
    v = v + 1;
end
subplot(3,3,7)
```

```
    %plot(acf{inp,7})
    axis([-0.5 10 -0.02 0.02])
    title('$<\phi_{n+l}X_n>$','Interpreter','latex')
    xlabel('time delay l')
    ylabel('$C_{20}(l)$','Interpreter','latex')
    hold on
    %% h) <phi_(n+l)alpha_n>
    v=1;
    for s=0:lag
        mult=0;
        summ=0;
        for t=1:m-s
            mult=((Phi{inp}(t+s)-ave3))*((Alpha{inp}(t)-ave2));
            summ=summ+mult;
        end
        acf{inp,8}(v,1)=summ/(m-s);
        v = v + 1;
    end
    subplot(3,3,8)
    %plot(acf{inp,8})
    axis([-0.5 10 -0.01 0.013])
    title('$<\phi_{n+l}\alpha_n>$','Interpreter','latex')
    xlabel('time delay l')
    ylabel('$C_{21}(l)$','Interpreter','latex')
    hold on
    %% i) <phi_(n+l)phi_n>
    v=1;
    for s=0:lag
        mult=0;
        summ=0;
        for t=1:m-s
            mult=((Phi{inp}(t+s)-ave3))*((Phi{inp}(t)-ave3));
            summ=summ+mult;
        end
        acf{inp,9}(v,1)=summ/(m-s);
        v = v + 1;
    end
    subplot(3,3,9)
    %plot(acf{inp,9})
    axis([-0.5 10 -0.03 0.03])
    title('$<\phi_{n+l}\phi_n>$','Interpreter','latex')
    xlabel('time delay l')
    ylabel('$C_{22}(l)$','Interpreter','latex')
    hold on
    clearvars -except source_dir source_files inp acf celllength celltime
data lag1 Alpha X Phi
end

%% calculate averages with error bars
for j = 1:9
    sum1 = 0;
    for i = 1:inp
        if ~isnan(acf{i,j})
            acf1 = acf{i,j};
            sum1 = sum1 + acf1;
        end
    end
```

```
    ave1 = sum1./inp;
    for e = 1:10
        l = 1;
        for i = 1:inp
            allpoints(e,l) = acf{i,j}(e,1);
            l = l + 1;
        end
    end
    zeroline = zeros(12,1);
    lag2(:,1) = linspace(-1,10,12);
    subplot(3,3,j)
    %plot(lag1, ave1,'b','linewidth',3)
    plot(lag2,zeroline,'k','linewidth',2)
    for i = 1:10
        err(i,1) = std(allpoints(i,:));
    end
    errorbar(lag1,ave1,err,'b-o')
end

%% Confined Autoregressive Model

for inp = 1:length(source_files)
    Y{inp} = [X{inp},Alpha{inp},Phi{inp}];
    for i = 1:size(X{inp})-1
        Y1{inp}(i,1) = X{inp}(i+1,1);
        Y1{inp}(i,2) = Alpha{inp}(i,1);
        Y1{inp}(i,3) = Phi{inp}(i,1);
        Y2{inp}(i,1) = X{inp}(i+1,1);
        Y2{inp}(i,2) = Alpha{inp}(i+1,1);
        Y2{inp}(i,3) = Phi{inp}(i,1);
    end
    Mdl{inp} = varm(3,3);
    EstMdl{inp} = estimate(Mdl{inp},Y{inp});
    EstMdl1{inp} = estimate(Mdl{inp},Y1{inp});
    EstMdl2{inp} = estimate(Mdl{inp},Y2{inp});

    %% R parameters for X
    ARparam{1,1}(inp,1) = 0;
    ARparam{1,1}(inp,2) = EstMdl{inp}.AR{1}(1,1);
    ARparam{1,1}(inp,3) = EstMdl{inp}.AR{2}(1,1);
    ARparam{1,2}(inp,1) = 0;
    ARparam{1,2}(inp,2) = EstMdl{inp}.AR{1}(1,2);
    ARparam{1,2}(inp,3) = EstMdl{inp}.AR{2}(1,2);
    ARparam{1,3}(inp,1) = 0;
    ARparam{1,3}(inp,2) = EstMdl{inp}.AR{1}(1,3);
    ARparam{1,3}(inp,3) = EstMdl{inp}.AR{2}(1,3);

    %% R parameters for alpha
    ARparam{2,1}(inp,1) = EstMdl1{inp}.AR{1}(2,1);
    ARparam{2,1}(inp,2) = EstMdl1{inp}.AR{2}(2,1);
    ARparam{2,1}(inp,3) = EstMdl1{inp}.AR{3}(2,1);
    ARparam{2,2}(inp,1) = 0;
    ARparam{2,2}(inp,2) = EstMdl1{inp}.AR{1}(2,2);
    ARparam{2,2}(inp,3) = EstMdl1{inp}.AR{2}(2,2);
    ARparam{2,3}(inp,1) = 0;
    ARparam{2,3}(inp,2) = EstMdl1{inp}.AR{1}(2,3);
```

```
    ARparam{2,3}(inp,3) = EstMdl1{inp}.AR{2}(2,3);

    %% R parameters for phi
    ARparam{3,1}(inp,1) = EstMdl2{inp}.AR{1}(3,1);
    ARparam{3,1}(inp,2) = EstMdl2{inp}.AR{2}(3,1);
    ARparam{3,1}(inp,3) = EstMdl2{inp}.AR{3}(3,1);
    ARparam{3,2}(inp,1) = EstMdl2{inp}.AR{1}(3,2);
    ARparam{3,2}(inp,2) = EstMdl2{inp}.AR{2}(3,2);
    ARparam{3,2}(inp,3) = EstMdl2{inp}.AR{3}(3,2);
    ARparam{3,3}(inp,1) = 0;
    ARparam{3,3}(inp,2) = EstMdl2{inp}.AR{1}(3,3);
    ARparam{3,3}(inp,3) = EstMdl2{inp}.AR{2}(3,3);

    %% constant values
    ARparam{1,4}(inp,1) = EstMdl{inp}.Constant(1,1);
    ARparam{2,4}(inp,1) = EstMdl1{inp}.Constant(1,1);
    ARparam{3,4}(inp,1) = EstMdl2{inp}.Constant(1,1);
end

%% Average values for R
figure(2)
k = 1;
num = [0;1;2];
zline = [0;0;0];
zxline = [-0.5;1;2.5];
for l = 1:3
    for j = 1:3
        for i = 1:3
            RR(k,i) = mean(ARparam{l,j}(:,i));
            errRR(k,i) = std(ARparam{l,j}(:,i));
        end
        subplot(3,3,k)
        errorbar(num,RR(k,:),errRR(k,:),'b-o','linewidth',2)
        hold on
        plot(zxline,zline,'k','linewidth',2)
        hold off
        axis([-0.5 2.5 -2 2])
        xlabel('time delay l')
        k = k + 1;
    end
end
subplot(3,3,1)
title('${R_{xx}}^{(l)}$','Interpreter','latex')
subplot(3,3,2)
title('${R_{x\alpha}}^{(l)}$','Interpreter','latex')
subplot(3,3,3)
title('${R_{x\phi}}^{(l)}$','Interpreter','latex')
subplot(3,3,4)
title('${R_{\alpha x}}^{(l)}$','Interpreter','latex')
subplot(3,3,5)
title('${R_{\alpha\alpha}}^{(l)}$','Interpreter','latex')
subplot(3,3,6)
title('${R_{\alpha\phi}}^{(l)}$','Interpreter','latex')
subplot(3,3,7)
title('${R_{\phi x}}^{(l)}$','Interpreter','latex')
subplot(3,3,8)
```

```
title('${R_{\phi\alpha}}^{(l)}$','Interpreter','latex')
subplot(3,3,9)
title('${R_{\phi\phi}}^{(l)}$','Interpreter','latex')
```

```
# BingKan Xue
# pool over multiple datasets, for each and both cells A, B
# infer autoregressive model using least square estimation

import sys, os, glob
import numpy as np
import matplotlib.pyplot as plt

p = 9    # order of autoregression

input_folders = [#'results6_042817', 'results5_101218',
                 'results6_062819',
                 'results3_081619', 'results4_092519']
output_folder = 'results_pool'
input_name = 'trace(crct)_'
output_name = 'lstsq%d_pool%d' % (p, len(input_folders))

files = []
for folder in input_folders:
    files.extend(glob.glob('%s/%s*.dat' % (folder, input_name)))
files.sort()

variables = [r'$\alpha_{n}$', r'$\tau_{n}$', r'$x_{n}$', r'$y_{n}$', r'$g_{n}$', r'$f_{n}$']
log_variables = [r'$\alpha_{n}$', r'$\tau_{n}$', r'$\chi_{n}$', r'$\eta_{n}$', r'$\phi_{n}$', r'$\psi_{n}$']
definitions = [r'$\alpha_{n}$', r'$\tau_{n}$', r'$x_{n}(0)$', r'$x_{n}(\tau_{n})$', r'$x_{n}(\tau_{n})/x_{n}(0)$', r'$x_{n}(0)/x_{n-1}(\tau_{n-1})$']
ranges = [[0.4,1.6], [0.2,1.4], [1.1,11.], [1.1,11.], [1.0,4.0], [0.3,0.7]]

log = True
if log:
    ranges = np.asarray(ranges)
    ranges[2:] = np.log(ranges[2:])
    variables = log_variables
    for i in range(2,len(variables)):
        definitions[i] = r'$\log\,%s' % definitions[i][1:]
vid = np.array([2, 0, 4])
var = [variables[i] for i in vid]
rng = [ranges[i] for i in vid]
info = 'variables: ' + ', '.join(var) + '\n'

lab = ['A', 'B', 'both']
col = plt.rcParams['axes.prop_cycle'].by_key()['color']

nv = len(vid)
nf = len(files)

num_all = []
dat_all = []
X_all = [[] for k in range(3)]    # for cell A, B, both
Y_all = [[] for k in range(3)]
est_all = [[] for k in range(3)]
cov_all = [[] for k in range(3)]
sig_all = [[] for k in range(3)]
phi_all = [[] for k in range(3)]
```

```
phi_err_all = [[] for k in range(3)]

# for each trace
for f in files:
    filename = os.path.basename(f)[len(input_name):-4]
    dat = np.loadtxt(f)
    if log:
        dat[:,2:] = np.log(dat[:,2:])
    num = dat.shape[0] / 2
    datA = dat[:num,vid]
    datB = dat[num:,vid]
    dat = np.hstack([datA, datB])
    if num < 20:    # minimum number of cycles
        continue
    dat_all.append(dat)
    num_all.append(num)
    print 'file: %s, #cycles: %d' % (filename, num)
    # estimate regression coefficients using least square
    Y = [None] * 3
    Y[0] = datA[p:,:]    # daughter cell A
    Y[1] = datB[p:,:]    # daughter cell B
    Y[2] = np.vstack((Y[0], Y[1]))    # both cells
    X = [None] * 3
    anc = np.hstack([datA[p-l:num-l,:] for l in range(1,p+1)]) # ancestor cell A
    X[0] = np.hstack([datA[p:,:], anc, np.ones((num-p,1))])
    X[1] = np.hstack([datB[p:,:], anc, np.ones((num-p,1))])
    X[2] = np.vstack((X[0], X[1]))
    for k in xrange(3):
        nanY = np.any(np.isnan(Y[k]), axis=1)
        nanX = np.any(np.isnan(X[k]), axis=1)
        ind = np.logical_and(~nanY, ~nanX)
        Y[k] = Y[k][ind]
        X[k] = X[k][ind]
        Y_all[k].append(Y[k])
        X_all[k].append(X[k])
    # estimate coefficients column by column
    est = [None] * 3
    cov = [None] * 3
    sig = [None] * 3
    phi = [None] * 3
    phi_err = [None] * 3
    cns = [None] * 3
    cns_err = [None] * 3
    for k in xrange(3):
        est[k] = np.zeros((X[k].shape[1], Y[k].shape[1]))
        for j in xrange(nv):
            nzj = np.array(range(j) + range(nv,X[k].shape[1]))
            est[k][nzj,j] = np.linalg.lstsq(X[k][:,nzj], Y[k][:,j])[0]
        est_all[k].append(est[k])
        # estimate noise covariance
        res = Y[k] - np.dot(X[k], est[k])    # residuals
        sig[k] = np.dot(res.T, res) / (Y[k].shape[0]-nv-1)
        sig_all[k].append(sig[k])
        # variance of estimation error for regression coefficients
        cov[k] = np.zeros((X[k].shape[1], Y[k].shape[1]))
        mat = np.dot(X[k].T, X[k])
        for j in xrange(nv):
            nzj = np.array(range(j) + range(nv,X[k].shape[1]))
            cov[k][nzj,j] = np.linalg.inv((mat[nzj,:])[:,nzj]).diagonal() * sig[k][j,j
    ]
        cov_all[k].append(cov[k])
```

```
        # transform to regression coefficients and constant terms
        cns[k] = est[k][-1,:]
        cns_err[k] = np.sqrt(cov[k][-1,:])
        phi[k] = np.reshape(est[k][:-1,:], (p+1,nv,nv)).swapaxes(1,2)
        phi_err[k] = np.reshape(np.sqrt(cov[k][:-1,:]), (p+1,nv,nv)).swapaxes(1,2)
        phi_all[k].append(phi[k])
        phi_err_all[k].append(phi_err[k])
    continue    # skip saving each trace
    # save results
    fout = open('%s/%s_%s.dat' % (output_folder, output_name, filename), 'w')
    fout.write('# for each cell A, B, and both cells:\n')
    for k in xrange(3):
        results = np.vstack((np.vstack(phi[k]), cns[k], sig[k], np.vstack(phi_err[k]),
 cns_err[k]))
        info_f = '\n' + info + 'number of cell cycles = %d\n' % num +\
                 'regression coefficients, constants, and noise variance (vstack):'
        np.savetxt(fout, results, fmt='%.8f', delimiter='\t', header=info_f)
    fout.close()
    # plot results
    fig, ax = plt.subplots(figsize=(3*(nv+1),3*nv), nrows=nv, ncols=nv+1)
    fig.subplots_adjust(left=0.07, right=0.92, bottom=0.07, top=0.92, wspace=0.5, hspa
ce=0.5)
    for i in xrange(nv):
      for j in xrange(nv):
        for k in xrange(3):
            ax[i,j].errorbar(np.arange(p+1), phi[k][:,i,j], yerr=phi_err[k][:,i,j], ma
rker='.', capsize=3, label=lab[k])
        ax[i,j].axhline(0, color='k', lw=1)
        ax[i,j].set_xlim(-0.5, 1.5)
        ax[i,j].set_xticks(np.arange(p+1))
        ax[i,j].set_xlabel(r'time delay $\ell$')
        ymax = 1.2
        ax[i,j].set_ylim(-ymax, ymax)
        ax[i,j].legend()
        ax[i,j].set_title(r'$R^{(\ell)}_{%s %s}$' % (var[i][1:-5], var[j][1:-5]))
    for i in xrange(nv):
        for k in xrange(3):
            ax[i,nv].errorbar([0], [cns[k][i]], yerr=[cns_err[k][i]], marker='.', caps
ize=2, label=lab[k], color=col[k])
            ax[i,nv].errorbar([1], [0], yerr=[np.sqrt(sig[k][i,i])], capsize=2, color=
col[k])
        ax[i,nv].axhline(0, color='k', lw=1)
        ax[i,nv].set_xlim(-0.5, 1.5)
        ax[i,nv].set_xticks([0, 1])
        ax[i,nv].set_xticklabels([r'$C$', r'$N$'])
        ax[i,nv].set_xlabel(r'constant, noise')
        ymax = 1.2
        ax[i,nv].set_ylim(-ymax, ymax)
        ax[i,nv].legend()
        ax[i,nv].set_title(r'$C_%s$, $N_%s$' % (var[i][1:-5], var[i][1:-5]))
    fig.suptitle('%d cycles' % num)
    plt.savefig('%s/%s_%s.png' % (output_folder, output_name, filename))
    plt.close(fig)

# for all traces
filename = 'all'
Y = [None] * 3
X = [None] * 3
for k in xrange(3):
    X[k] = np.vstack(X_all[k])
    Y[k] = np.vstack(Y_all[k])
```

```
num = X[0].shape[0]
print 'file: %s, #cycles: %d' % (filename, num)
# estimate regression coefficients using least square
est = [None] * 3
cov = [None] * 3
sig = [None] * 3
phi = [None] * 3
phi_err = [None] * 3
cns = [None] * 3
cns_err = [None] * 3
for k in xrange(3):
    est[k] = np.zeros((X[k].shape[1], Y[k].shape[1]))
    for j in xrange(nv):    # estimate column by column
        nzj = np.array(range(j) + range(nv,X[k].shape[1]))
        est[k][nzj,j] = np.linalg.lstsq(X[k][:,nzj], Y[k][:,j])[0]
    # estimate noise covariance
    res = Y[k] - np.dot(X[k], est[k])    # residuals
    sig[k] = np.dot(res.T, res) / (Y[k].shape[0]-nv-1)
    # variance of estimation error for regression coefficients
    cov[k] = np.zeros((X[k].shape[1], Y[k].shape[1]))
    mat = np.dot(X[k].T, X[k])
    for j in xrange(nv):
        nzj = np.array(range(j) + range(nv,X[k].shape[1]))
        cov[k][nzj,j] = np.linalg.inv((mat[nzj,:])[:,nzj]).diagonal() * sig[k][j,j]
    # transform to regression coefficients and constant terms
    cns[k] = est[k][-1,:]
    cns_err[k] = np.sqrt(cov[k][-1,:])
    phi[k] = np.reshape(est[k][:-1,:], (p+1,nv,nv)).swapaxes(1,2)
    phi_err[k] = np.reshape(np.sqrt(cov[k][:-1,:]), (p+1,nv,nv)).swapaxes(1,2)
# save results
fout = open('%s/%s_%s.dat' % (output_folder, output_name, filename), 'w')
fout.write('# datasets: %s\n\n' % input_folders)
for k in xrange(3):
    fout.write('# for cell %s:\n' % lab[k])
    results = np.vstack((np.vstack(phi[k]), cns[k], sig[k], np.vstack(phi_err[k]), cns
_err[k]))
    info_f = '\n' + info + 'number of cell cycles = %d\n' % num +\
             'regression coefficients, constants, and noise variance (vstack):'
    np.savetxt(fout, results, fmt='%.8f', delimiter='\t', header=info_f)
fout.close()
# plot results
phi = np.array(phi)
phi_err = np.array(phi_err)
cns = np.array(cns)
cns_err = np.array(cns_err)
sig = np.array(sig)
fig, ax = plt.subplots(figsize=(3*(nv+1),3*nv), nrows=nv, ncols=nv+1)
fig.subplots_adjust(left=0.07, right=0.92, bottom=0.07, top=0.92, wspace=0.5, hspace=0
.5)
for i in xrange(nv):
  for j in xrange(nv):
    for k in xrange(3):
        ax[i,j].errorbar(np.arange(p+1), phi[k,:,i,j], yerr=phi_err[k,:,i,j],
                         marker='.', capsize=3, label=lab[k])
    ax[i,j].axhline(0, color='k', lw=1)
    ax[i,j].set_xlim(-0.5, p+0.5)
    ax[i,j].set_xticks(np.arange(p+1))
    ax[i,j].set_xlabel(r'time delay $\ell$')
#    ymax = 1.2 * np.amax(np.abs(phi[:,:,i,j]) + phi_err[:,:,i,j])
    ymax = 1.2
    ax[i,j].set_ylim(-ymax, ymax)
```

```
    ax[i,j].legend()
    ax[i,j].set_title(r'$R^{(\ell)}_{%s %s}$' % (var[i][1:-5], var[j][1:-5]))
for i in xrange(nv):
    for k in xrange(3):
        ax[i,nv].errorbar([0], [cns[k,i]], yerr=[cns_err[k,i]], marker='.', capsize=2, label=lab[k], color=col[k])
        ax[i,nv].errorbar([1], [0], yerr=[np.sqrt(sig[k,i,i])], capsize=2, color=col[k])
    ax[i,nv].axhline(0, color='k', lw=1)
    ax[i,nv].set_xlim(-0.5, 1.5)
    ax[i,nv].set_xticks([0, 1])
    ax[i,nv].set_xticklabels([r'$C$', r'$N$'])
    ax[i,nv].set_xlabel(r'constant, noise')
#    ymax = 1.2 * np.amax([np.abs(cns[:,i])+cns_err[:,i], np.sqrt(sig[:,i,i])])
    ymax = 1.2
    ax[i,nv].set_ylim(-ymax, ymax)
    ax[i,nv].legend()
    ax[i,nv].set_title(r'$C_%s$, $N_%s$' % (var[i][1:-5], var[i][1:-5]))
fig.suptitle('%d cycles' % num)
plt.savefig('%s/%s_%s.png' % (output_folder, output_name, filename))
plt.close(fig)

# average over traces
filename = 'avg'
nf = len(X_all[0])
print 'file: %s, #traces: %d' % (filename, nf)
est_mean = [None] * 3
est_se = [None] * 3
cov = [None] * 3
sig_mean = [None] * 3
phi_mean = [None] * 3
phi_se = [None] * 3
cns_mean = [None] * 3
cns_se = [None] * 3
for k in xrange(3):
    sig_mean[k] = np.mean(sig_all[k], axis=0)
    est_mean[k] = np.mean(est_all[k], axis=0)
    est_se[k] = np.std(est_all[k], axis=0) / np.sqrt(nf)
    # transform to regression coefficients and constant terms
    cns_mean[k] = est_mean[k][-1,:]
    cns_se[k] = est_se[k][-1,:]
    phi_mean[k] = np.reshape(est_mean[k][:-1,:], (p+1,nv,nv)).swapaxes(1,2)
    phi_se[k] = np.reshape(est_se[k][:-1,:], (p+1,nv,nv)).swapaxes(1,2)
# save results
fout = open('%s/%s_%s.dat' % (output_folder, output_name, filename), 'w')
fout.write('# datasets: %s\n\n' % input_folders)
for k in xrange(3):
    fout.write('# for cell %s:\n' % lab[k])
    results = np.vstack((np.vstack(phi_mean[k]), cns_mean[k], sig_mean[k], np.vstack(phi_se[k]), cns_se[k]))
    info_f = '\n' + info + 'number of cell cycles = %d\n' % num +\
             'regression coefficients, constants, and noise variance (vstack):'
    np.savetxt(fout, results, fmt='%.8f', delimiter='\t', header=info_f)
fout.close()
# plot results
fig, ax = plt.subplots(figsize=(3*(nv+1),3*nv), nrows=nv, ncols=nv+1)
fig.subplots_adjust(left=0.07, right=0.92, bottom=0.07, top=0.92, wspace=0.5, hspace=0.5)
for i in xrange(nv):
  for j in xrange(nv):
    for k in xrange(3):
```

```
286.         ax[i,j].errorbar(np.arange(p+1), phi_mean[k][:,i,j], yerr=phi_se[k][:,i,j], ma
    rker='.', capsize=3, label=lab[k])
287.     ax[i,j].axhline(0, color='k', lw=1)
288.     ax[i,j].set_xlim(-0.5, p+0.5)
289.     ax[i,j].set_xticks(np.arange(p+1))
290.     ax[i,j].set_xlabel(r'time delay $\ell$')
291.#     ymax = 1.2 * np.amax(np.abs(phi[:,:,i,j]) + phi_err[:,:,i,j])
292.     ymax = 1.2
293.     ax[i,j].set_ylim(-ymax, ymax)
294.     ax[i,j].legend()
295.     ax[i,j].set_title(r'$R^{(\ell)}_{%s %s}$' % (var[i][1:-5], var[j][1:-5]))
296.for i in xrange(nv):
297.     for k in xrange(3):
298.         ax[i,nv].errorbar([0], [cns_mean[k][i]], yerr=[cns_se[k][i]], marker='.', caps
    ize=2, label=lab[k], color=col[k])
299.         ax[i,nv].errorbar([1], [0], yerr=[np.sqrt(sig_mean[k][i,i])], capsize=2, color
    =col[k])
300.     ax[i,nv].axhline(0, color='k', lw=1)
301.     ax[i,nv].set_xlim(-0.5, 1.5)
302.     ax[i,nv].set_xticks([0, 1])
303.     ax[i,nv].set_xticklabels([r'$C$', r'$N$'])
304.     ax[i,nv].set_xlabel(r'constant, noise')
305.#     ymax = 1.2 * np.amax([np.abs(cns[:,i])+cns_err[:,i], np.sqrt(sig[:,i,i])])
306.     ymax = 1.2
307.     ax[i,nv].set_ylim(-ymax, ymax)
308.     ax[i,nv].legend()
309.     ax[i,nv].set_title(r'$C_%s$, $N_%s$' % (var[i][1:-5], var[i][1:-5]))
310.fig.suptitle('%d cycles' % num)
311.plt.savefig('%s/%s_%s.png' % (output_folder, output_name, filename))
312.plt.close(fig)
```

Appendix C Simulating null model

To test our methods for extracting cell cycle variables from data and doing multi-regression analysis in Section 3.3, we simulated cell growth and length measurement using a null model. This model is based on the "adder" mechanism and does not include any other correlation between cell cycle variables within or across generations. By applying our methods on the simulated data, we were able to estimate the systematic errors and obtain a baseline for the inferred regression coefficients.

Let x_n^a and y_n^a be the initial and final sizes of cell $a = A,$ or B in the n-th cell cycle. The null model asserts that $y_n^a = x_n^a + \Delta$, where Δ is a random variable that is drawn independently for each cell cycle. To simulate the null model, we used the empirical distribution of Δ, denoted by $P(\Delta)$, the empirical distribution of the growth rate, $P(\alpha)$, and the empirical distribution of the division fraction, $P(f)$. Those distributions are collected from the measured cell cycle variables (using curve-fitting as described above).

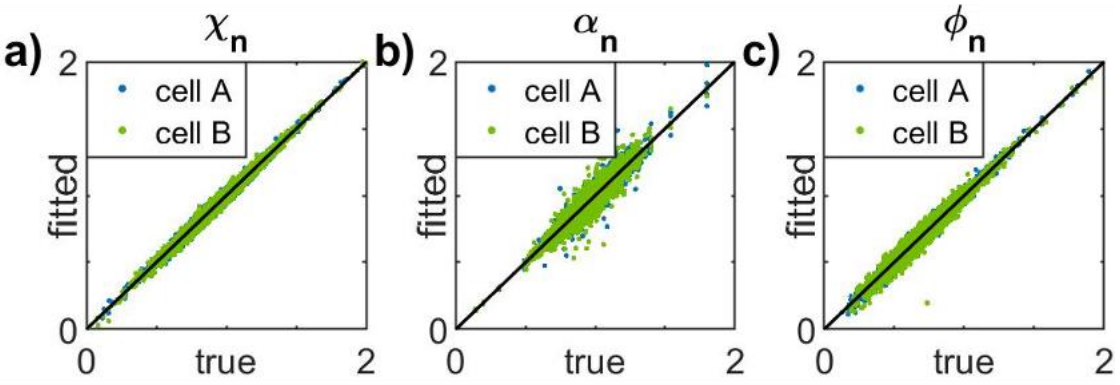
A cell lineage of $(N + 2)$ cell cycles is simulated as follows (two extra cycles are added because later when we fit growth curves, the first and last cycles are always treated as incomplete and discarded). Starting from a size y_0^A (e.g., 2Δ), randomly draw a fraction f_1^A from $P(f)$ and let $x_1^A = f_1^A\, y_0^A$ and $x_1^B = (1 - f_1^A) y_0^A$. Then randomly draw two numbers Δ_1^A and Δ_1^B from $P(\Delta)$, and let $y_1^A = x_1^A + \Delta_1^A$ and $y_1^B = x_1^B + \Delta_1^B$. Also draw two numbers α_1^A and α_1^B from $P(\alpha)$. Those numbers completely determine the growth curves of cell A and B in the 1-st generation. The same procedure is repeated for the next generation, starting from y_1^A, and so on. The other cell cycle

variables can be calculated from those numbers, such as $\phi_n^a = \log(y_n^a/x_n^a)$ and $\tau_n^a = \phi_n^a/\alpha_n^a$. We will refer to these simulated cell cycle variables as their "true" values.

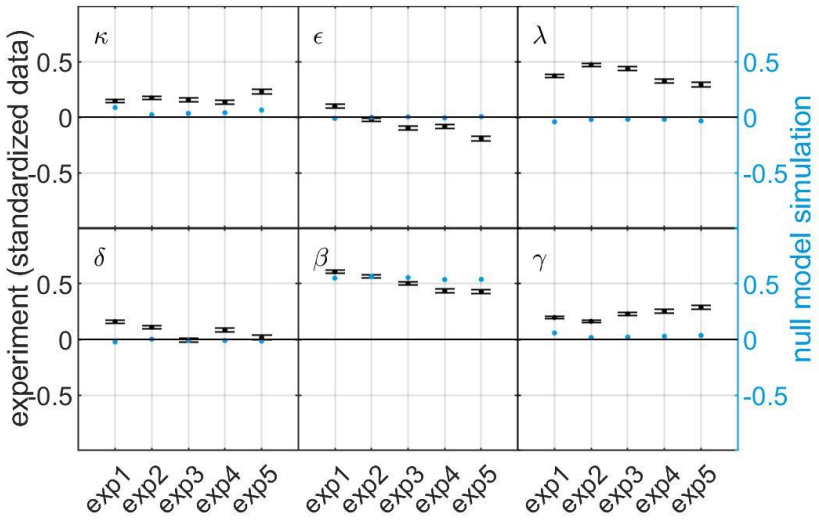
To mimic the measurement process and error, we reconstructed the cell length as a function of time, using the "true" values of cell cycle variables and assuming exponential growth. Consecutive cell cycles of cell A are concatenated continuously in time, whereas the beginning time of cell B is aligned with that of cell A in each generation. Then we sampled the reconstructed time courses of cell lengths at every 3 min. On top of these "true" cell lengths, we added a random noise to each data point, which is independently drawn from a Gaussian distribution with mean zero and variance σ^2 (obtained from curve-fitting, as described above). These noise-added discrete-time data points are our "measured" cell lengths, which we used as simulated data.

Finally, we used the same methods as for the real data to fit growth curves and extract cell cycle variables from the simulated data. We will refer to these as the "fitted" values of the cell cycle variables, which can be compared to their "true" values. Such a comparison is shown in Appendix Figure 1, and can be used to estimate the systematic errors incurred during the curve fitting.

Since in the simulation the "true" cell cycle variables are drawn independently, it is expected that all regression coefficients, except $\beta = 0.5$ which comes from the adder mechanism, should be zero. Therefore, any other coefficient found in the "fitted" cell cycle variables would be a systematic error of the curve-fitting and multi-regression methods. We evaluated the regression coefficients from such "fitted" cell cycle variables and used their values as a "baseline" for judging the significance of the coefficients found in the real data, as shown with blue dots in Figure 20 and Appendix Figure 2.

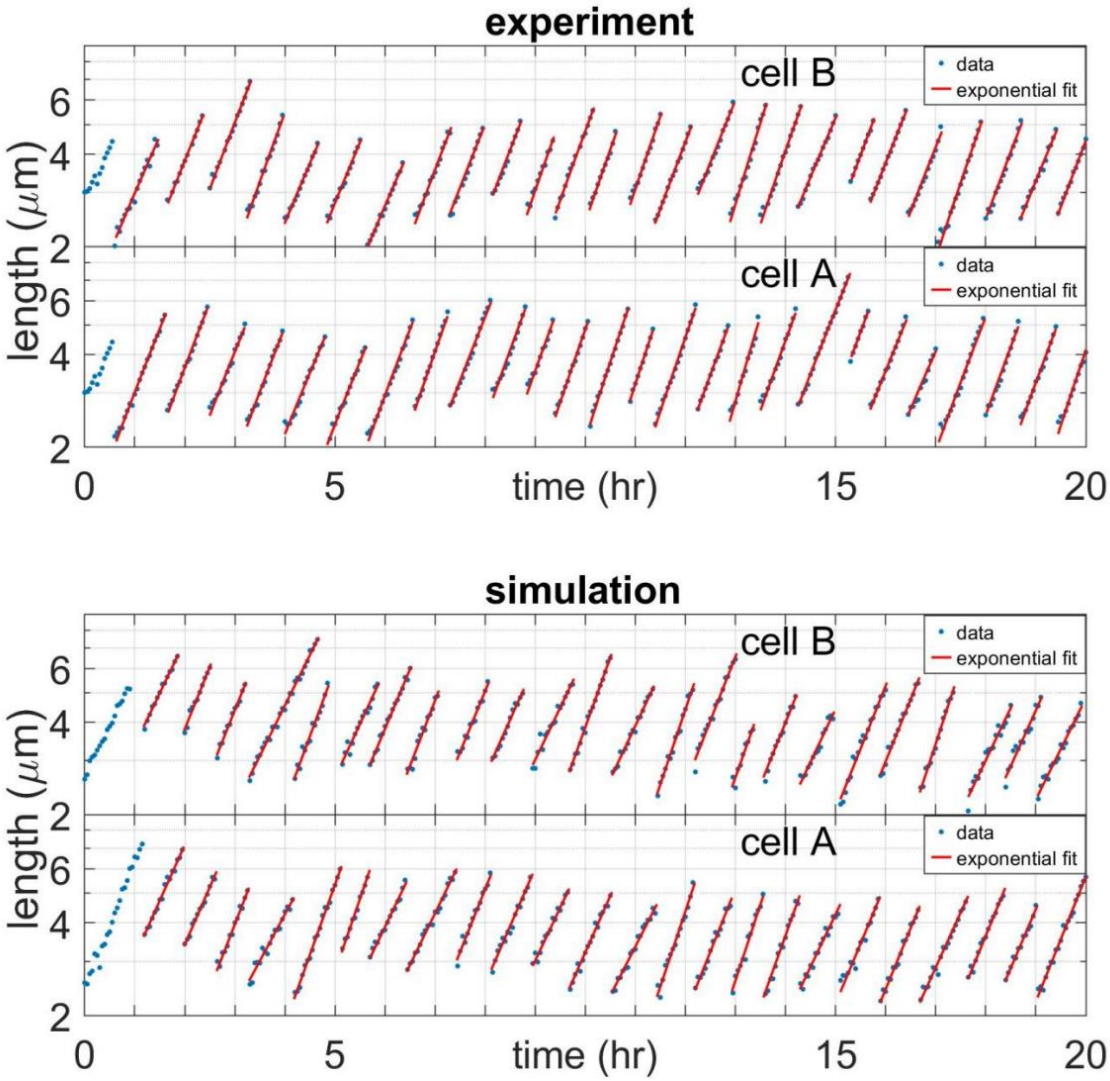


Appendix Figure 1 Simulation results of a null model based on the "adder" mechanism. The adder model is used to simulate the dynamics of cell cycle variables over many generations. Three such variables are shown here, the log initial size χ_n, the growth rate α_n, and the log relative growth ϕ_n. Their "true" values are those coming from the simulation, and their "fitted" values are those given by our curve-fitting method applied to simulated noisy measurements. The difference between "true" and "fitted" values shows the extent of systematic error in our curve-fitting method.



Appendix Figure 2 Comparison of estimated parameters across experiments with their baselines after standardizing cell cycle variables. The same data as for Figure 20 is used to fit the parameters, except that each cell cycle variable is standardized (i.e., removed the mean and normalized by the standard deviation) before fitting. Like in Figure 20, blue dots represent the baselines for the parameters found using simulated

data from the null model, with cell cycle variables also standardized before fitting . Values of ε and δ are negligible and fluctuate around zero.



Appendix Figure 3 Cell size growth curves fitting and null model simulation. Examples of cell size dynamics for a pair of sister cells from our experiments are presented for comparison with simulation results. In the experimental data, cell A denotes the cell at the end of a growth channel, and cell B is its sister cell. The simulations were carried out using the adder model, as described above. The red lines depict exponential growth curves during individual cell cycles, fitted using the procedure described in Appendix A. The experimental data are the same as that used in Figure 16 of the main text. The simulation results are used to estimate the systematic errors of our data analysis methods.

Algorithm for simulating null model:

```
# BingKan Xue
# resample cell traces by making simulated data
# use ADDER mechanism to stabilize size distribution
# randomly draw growth rate, final size, division fraction for each cell
# using empirical distribution of each variable

import sys, os, glob
import numpy as np
import matplotlib.pyplot as plt

#model = 'timer'
#model = 'sizer'
model = 'adder'

folder = 'results6_042817'
input_name = '042817'
output_name = model
output_name1 = 'trace(sim)'

variables = [r'$\alpha_{n}$', r'$\tau_{n}$', r'$x_{n}$', r'$y_{n}$', r'$g_{n}$', r'$f_{n}$']
log_variables = [r'$\alpha_{n}$', r'$\tau_{n}$', r'$\chi_{n}$', r'$\eta_{n}$', r'$\phi_{n}$', r'$\psi_{n}$']
definitions = [r'$\alpha_{n}$', r'$\tau_{n}$', r'$x_{n}(0)$', r'$x_{n}(\tau_{n})$', r'$x_{n}(\tau_{n})/x_{n}(0)$', r'$x_{n}(0)/x_{n-1}(\tau_{n-1})$']
ranges = [[0.4,1.6], [0.2,1.4], [1.1,11.], [1.1,11.], [1.0,4.0], [0.3,0.7]]

# load number and lengths of traces
hist = np.loadtxt('%s/%s.dat' % (folder, 'extract(crct)_hist'), dtype=int)
num_all = np.repeat(hist[:,0], hist[:,1])

# read empirical distribution of variables
dist = np.load('%s/%s.npz' % (folder, 'dist_all'))
dat_all = dist['dat_all']
nv = len(variables)
datA = dat_all[:,:nv]
datB = dat_all[:,nv:]
dat = np.vstack((datA, datB))
emp = []
for i in range(nv):
    ind = np.isnan(dat[:,i])
    emp.append(dat[~ind,i])
if model == 'adder':
    add = dat[:,3] - dat[:,2]
    ind = np.isnan(add)
    add = add[~ind]

# read std of residuals
dist = open('%s/%s.dat' % (folder, 'dist_all'), 'r')
for line in dist:
    if line.startswith('std of residuals', 2):
        i = line.index('=')
        resid = float(line[i+2:-1])
        break

# simulate data
nf = len(num_all)    # number of traces
```

```
num_all = np.random.choice(num_all, size=10*nf, replace=True)
for n in xrange(10*nf):
    filename = '%s%d' % (output_name, n+1)
    num = num_all[n]
    print 'file: %s, #cycles: %d' % (filename, num)

    # draw variables that determine each cycle (first and last will not count)
    rateA = np.random.choice(emp[0], size=(num+2))    # resample empirical dist
    rateB = np.random.choice(emp[0], size=(num+2))
    fracA = np.random.choice(emp[5], size=(num+2))
    fracB = 1. - fracA
    initA = np.zeros(num+2)
    initB = np.zeros(num+2)
    if model == 'timer':
        duraA = np.random.choice(emp[1], size=(num+2))
        duraB = np.random.choice(emp[1], size=(num+2))
        finaA = np.mean(emp[3])
        for i in xrange(num+2):
            initA[i] = finaA * fracA[i]
            initB[i] = finaA * fracB[i]
            finaA = initA[i] * np.exp(rateA[i] * duraA[i])    # timer
    elif model == 'sizer':
        duraA = np.zeros(num+2)
        duraB = np.zeros(num+2)
        finaA = np.random.choice(emp[3], size=(num+2))
        finaB = np.random.choice(emp[3], size=(num+2))
        for i in xrange(num+2):
            initA[i] = finaA[i-1] * fracA[i]  # first cycle uses last final size
            initB[i] = finaA[i-1] * fracB[i]
            if finaA[i] <= initA[i] or finaB[i] <= initB[i]:
                raise RuntimeError('size decreased!')
            duraA[i] = np.log(finaA[i] / initA[i]) / rateA[i]    # sizer
            duraB[i] = np.log(finaB[i] / initB[i]) / rateB[i]
    elif model == 'adder':
        duraA = np.zeros(num+2)
        duraB = np.zeros(num+2)
        addA = np.random.choice(add, size=(num+2))
        addB = np.random.choice(add, size=(num+2))
        finaA = np.mean(emp[3])
        for i in xrange(num+2):
            initA[i] = finaA * fracA[i]
            initB[i] = finaA * fracB[i]
            finaA = initA[i] + addA[i]    # adder
            finaB = initB[i] + addB[i]
            duraA[i] = np.log(finaA / initA[i]) / rateA[i]
            duraB[i] = np.log(finaB / initB[i]) / rateB[i]
    start = np.cumsum(duraA) - duraA
    if np.any(duraA < 0.2) or np.any(duraB < 0.2):
        print 'cycle too short'
        continue
    # simulate time series of cell length
    dt = 1./12
    time = np.arange(0., start[-1] + duraA[-1], dt)
    timeA, sizeA = [], []
    timeB, sizeB = [], []
    for i in xrange(num+2):
        beg = np.nonzero(time >= start[i])[0][0]
        endA = np.nonzero(time < start[i] + duraA[i])[0][-1]
        indA = np.arange(beg, endA+1)
        if len(indA) < 2:
            raise RuntimeError('cycle A too short')
```

```
        tA = time[indA]
        sA = initA[i] * np.exp(rateA[i] * (tA - start[i]))
        timeA.append(tA.copy())
        sizeA.append(sA.copy())
        endB = np.nonzero(time < start[i] + duraB[i])[0][-1]
        indB = np.arange(beg, endB+1)
        if len(indB) < 2:
            raise RuntimeError('cycle B too short')
        tB = time[indB]
        sB = initB[i] * np.exp(rateB[i] * (tB - start[i]))
        timeB.append(tB.copy())
        sizeB.append(sB.copy())
    if len(timeA) != len(timeB):
        raise RuntimeError('number of cycles not match')
    elif not np.allclose([tA[0] for tA in timeA], [tB[0] for tB in timeB]):
        raise RuntimeError('beginning times not match')
    timeA = np.hstack(timeA)
    sizeA = np.hstack(sizeA)
    timeB = np.hstack(timeB)
    sizeB = np.hstack(sizeB)
    # add measurement noise
    sizeA += resid * np.random.randn(len(sizeA))
    sizeB += resid * np.random.randn(len(sizeB))
    # save simulated data points
    dat = np.vstack([np.hstack([timeA, timeB]), np.hstack([sizeA, sizeB])]).T
    np.savetxt('%s/%s_%s.csv' % (folder, input_name, filename), dat,
               fmt='%.6f', delimiter=', ')

    # save true cell cycle variables
    slopeA = rateA[1:-1]
    duratA = duraA[1:-1]
    initA = initA[1:-1]
    growA = np.exp(rateA[1:-1] * duraA[1:-1])
    finaA = initA * growA
    fracA = fracA[1:-1]
    slopeB = rateB[1:-1]
    duratB = duraB[1:-1]
    initB = initB[1:-1]
    growB = np.exp(rateB[1:-1] * duraB[1:-1])
    finaB = initB * growB
    fracB = fracB[1:-1]
    fout = open('%s/%s_%s.dat' % (folder, output_name1, filename), 'w')
    fout.write('# exponential fit of cell length vs time\n' +\
               '# standard deviation of residuals = %.6f\n' % resid)
    info = 'slope, duration, initial size, final size, ' +\
           'growth factor, division fraction:'
    info += '\ncell A (mother):'
    results = np.vstack((slopeA, duratA, initA, finaA, growA, fracA)).T
    np.savetxt(fout, results, fmt = '%.8f', delimiter='\t', header=info)
    info = 'cell B (daughter):'
    results = np.vstack((slopeB, duratB, initB, finaB, growB, fracB)).T
    np.savetxt(fout, results, fmt = '%.8f', delimiter='\t', header=info)
    fout.close()
```

Appendix D List of mutants

Appendix Table 1 List of isogenic single-gene deletion *E. coli* K12 BW25113 mutants obtained from the Keio collection.[55,136,137,138]

Gene Name	location, protein, subunit
aspA	cytosol, aspartate ammonia-lyase
atpA	inner membrane, ATP synthase; F1 sector; alpha-subunit
atpB	inner membrane, ATP synthase; F0 sector; subunit a
atpC	inner membrane, ATP synthase; F1 sector; epsilon-subunit
atpD	inner membrane, ATP synthase; F1 sector; beta-subunit
atpE	inner membrane, ATP synthase; F0 sector; subunit c
atpF	inner membrane, ATP synthase; F0 sector; subunit b
atpG	inner membrane, ATP synthase; F1 sector; gamma-subunit
atpH	inner membrane, ATP synthase; F1 sector; delta-subunit
cpxR	cytosol, DNA-binding transcriptional dual regulator CpxR
cydD	inner membrane, glutathione/L-cysteine ABC exporter subunit CydD
cysB	cytosol, DNA-binding transcriptional dual regulator CysB
cysC	cytosol, adenylyl-sulfate kinase
cysD	cytosol, sulfate adenylyltransferase subunit 2
cysE	cytosol, serine acetyltransferase
cysG	cytosol, siroheme synthase
cysH	cytosol, phosphoadenosine phosphosulfate reductase
cysI	sulfite reductase, hemoprotein subunit
cysN	cytosol, sulfate adenylyltransferase subunit 1
cysP	periplasmic space, hiosulfate/sulfate ABC transporter periplasmic binding protein CysP
cysU	inner membrane, sulfate/thiosulfate ABC transporter inner membrane subunit CysU
cysW	inner membrane, sulfate/thiosulfate ABC transporter inner membrane subunit CysW
degP	inner membrane, periplasmic space, periplasmic serine endoprotease DegP
fkpB	cytosol, peptidyl-prolyl cis-trans isomerase FkpB
folB	cytosol, dihydroneopterin aldolase
gmhB	cytosol, D-sedoheptulose 7-phosphate isomerase; GDP-heptose biosynthesis; T-phage resistance
guaA	cytosol, GMP synthetase
hyaD	cytosol, putative hydrogenase 1 maturation protease HyaD
iscS	cytosol, cysteine desulfurase

lipA	cytosol, lipoyl synthase
lipB	cytosol, lipoyl(octanoyl) transferase
lpd	inner membrane, cytosol, lipoamide dehydrogenase
nlpI	inner membrane, outer membrane, lipoprotein NlpI
nuoA	inner membrane, NADH:quinone oxidoreductase subunit A
nuoB	inner membrane, NADH:quinone oxidoreductase subunit B
nuoC	inner membrane, NADH:quinone oxidoreductase subunit CD
nuoE	inner membrane, NADH:quinone oxidoreductase subunit E
nuoF	inner membrane, NADH:quinone oxidoreductase subunit F
nuoG	inner membrane, NADH:quinone oxidoreductase subunit G
nuoH	inner membrane, NADH:quinone oxidoreductase subunit H
nuoI	inner membrane, NADH:quinone oxidoreductase subunit I
nuoJ	inner membrane, NADH:quinone oxidoreductase subunit J
nuoK	inner membrane, NADH:quinone oxidoreductase subunit K
nuoL	inner membrane, NADH:quinone oxidoreductase subunit L
nuoM	inner membrane, NADH:quinone oxidoreductase subunit M
nuoN	inner membrane, NADH:quinone oxidoreductase subunit N
prc	periplasmic space, inner membrane, tail-specific protease
recB	cytosol, exodeoxyribonuclease V subunit RecB
rimM	cytosol, ribosome, ribosome maturation factor RimM
rseA	inner membrane, anti-sigma-E factor RseA
sdhB	inner membrane, succinate:quinone oxidoreductase, iron-sulfur cluster binding protein
sucA	cytosol, dihydrolipoyltranssuccinylase
sucB	cytosol, dihydrolipoyltranssuccinylase
tolA	inner membrane, Tol-Pal system protein TolA
tolB	periplasmic space,Tol-Pal system periplasmic protein TolB
tolR	inner membrane, Tol-Pal system protein TolR
tpiA	cytosol, triose-phosphate isomerase
ubiE	bifunctional 2-octaprenyl-6-methoxy-1,4-benzoquinone methylase and S-adenosylmethionine:2-DMK methyltransferase
ubiF	cytosol, 2-octaprenyl-3-methyl-6-methoxy-1,4-benzoquinol hydroxylase
ubiG	inner membrane, cytosol, bifunctional 3-demethylubiquinone-8 3-O-methyltransferase and 2-octaprenyl-6-hydroxyphenol methylase
ubiH	cytosol, 2-octaprenyl-6-methoxyphenol to 2-octaprenyl-6-methoxy-1;4-benzoquinone
yciH	cytosol, putative translation factor
ydjI	cytosol, putative aldolase YdjI
yeeS	cytosol, CP4-44 prophage; RadC-like JAB domain-containing protein YeeS

Appendix E Individual mother machine experiments for death phenotype statistics

Appendix Table 2 Death phenotypes of individual Δ*atpA* and *wt* cells grown in the indicated growth media (LB, M9CG or M9G). Numbers (exp. nom.) indicate individual experiments.

		exp. nom.	experiment time (hrs)	tot. num. of cells	dead cells %	Type I cells %	Type II cells %	Type III cells %
wild type	LB	1	122	84	77.4	20.4	59.2	20.4
	M9CG	1	161	68	97.0	14.3	60.7	25.0
		2	104	629	96.0	17.5	75.7	6.8
	M9G	1	88	60	84.0	10.6	19.2	70.2
		2	50	84	55.0	10.5	18.4	71.1
Δ*atpA*	LB	1	97	91	78.3	0	73.3	26.7
		2	102	28	85.7	0	79.2	20.8
	M9CG	1	140	99	98	0	81.5	18.5
		2	151	42	100	0	85.7	14.3
	M9G	1	160	40	100	0	56.4	43.6

Bibliography

1. Roepke, T. A. *et al.* Contribution of a membrane estrogen receptor to the estrogenic regulation of body temperature and energy homeostasis. *Endocrinology* **151**, 4926–4937 (2010).

2. Röder, P. V., Wu, B., Liu, Y. & Han, W. Pancreatic regulation of glucose homeostasis. *Exp. Mol. Med.* **48**, e219 (2016).

3. Guyenet, P. G. & Bayliss, D. A. Neural Control of Breathing and CO2 Homeostasis. *Neuron* **87**, 946–961 (2015).

4. Mitchison, J. M. Single cell studies of the cell cycle and some models. *Theor. Biol. Med. Model.* **2**, 1–5 (2005).

5. Zhang, Qing;Brambilla, Elisa;Li, Rui;Shi, Haulin;Lagomarsino, Marco Cosentino;Sclavi, B. A Decrease in Transcription Capacity Limits Growth Rate upon Translation Inhibition. *mSystems* **5**, 1–15 (2020).

6. Spivey, E. C., Jones, S. K., Rybarski, J. R., Saifuddin, F. A. & Finkelstein, I. J. An aging-independent replicative lifespan in a symmetrically dividing eukaryote. *Elife* **6**, 1–25 (2017).

7. Rang, C. U., Peng, A. Y., Poon, A. F. & Chao, L. Ageing in Escherichia coli requires damage by an extrinsic agent. *Microbiol. (United Kingdom)* **158**, 1553–1559 (2012).

8. Facchetti, G., Chang, F. & Howard, M. Controlling cell size through sizer mechanisms. *Curr. Opin. Syst. Biol.* **5**, 86–92 (2017).

9. Deforet, M., Van Ditmarsch, D. & Xavier, J. B. Cell-Size Homeostasis and the Incremental Rule in a Bacterial Pathogen. *Biophys. J.* **109**, 521–528 (2015).

10. Taheri-Araghi, S. *et al.* Cell-size control and homeostasis in bacteria. *Curr. Biol.* **25**, 385–391 (2015).

11. Si, F. *et al.* Mechanistic Origin of Cell-Size Control and Homeostasis in Bacteria. *Curr. Biol.* **29**, 1760-1770.e7 (2019).

12. Amir, A. Cell size regulation in bacteria. *Phys. Rev. Lett.* **112**, 1–5 (2014).

13. Susman, L. *et al.* Individuality and slow dynamics in bacterial growth homeostasis. *Proc. Natl. Acad. Sci. U. S. A.* **115**, (2018).

14. Ueno, T., Taguchi, H., Tadakuma, H., Yoshida, M. & Funatsu, T. GroEL mediates protein folding with a two successive timer mechanism. *Mol. Cell* **14**, 423–434 (2004).

15. Micali, G., Grilli, J., Marchi, J., Osella, M. & Cosentino Lagomarsino, M. Dissecting the Control Mechanisms for DNA Replication and Cell Division in E. coli. *Cell Rep.* **25**, 761-771.e4 (2018).

16. Osella, M., Tans, S. J. & Cosentino Lagomarsino, M. Step by Step, Cell by Cell: Quantification of the Bacterial Cell Cycle. *Trends Microbiol.* **25**, 250–256 (2017).

17. Nordholt, N., van Heerden, J. H. & Bruggeman, F. J. Biphasic Cell-Size and Growth-Rate Homeostasis by Single Bacillus subtilis Cells. *Curr. Biol.* **30**, 2238-2247.e5 (2020).

18. Salman, H. Bacterial Growth: Cell-Cycle Dependent Growth-Rate Homeostasis. *Curr. Biol.* **30**, R703–R704 (2020).

19. Männik, J., Woldringh, C. L. & Zaritsky, A. Editorial: The Bacterial Cell: Coupling between Growth, Nucleoid Replication, Cell Division, and Shape. *Front. Microbiol.* **7**, 116 (2016).

20. Amir, A., Männik, J., Woldringh, C. L. & Zaritsky, A. Editorial: The Bacterial Cell: Coupling between Growth, Nucleoid Replication, Cell Division, and Shape Volume 2. *Front. Microbiol.* **10**, 2056 (2019).

21. Smith, F. A., Lyons, S. K., Smith, F. A. & Lyons, S. K. On Being the Right Size. *Anim. Body Size* **348**, 1–10 (2015).

22. Ginzberg, M. B. *et al.* Cell size sensing in animal cells coordinates anabolic growth rates and cell cycle progression to maintain cell size uniformity. *Elife* **7**, 1–27 (2018).

23. Feijó Delgado, F. *et al.* Intracellular Water Exchange for Measuring the Dry Mass, Water Mass and Changes in Chemical Composition of Living Cells. *PLoS One* **8**, (2013).

24. Dennis, P. P. & Bremer, H. *Modulation of Chemical Composition and Other Parameters of the Cell at Different Exponential Growth Rates*. *Escherichia coli and Salmonella: cellular and molecular biology* vol. 2.2 (1996).

25. Cox, R. A. Correlation of the rate of protein synthesis and the third power of the RNA: Protein ratio in Escherichia coli and Mycobacterium tuberculosis. *Microbiology* **149**, 729–737 (2003).

26. Huh, D. & Paulsson, J. Random partitioning of molecules at cell division. *Proc. Natl. Acad. Sci. U. S. A.* **108**, 15004–15009 (2011).

27. Kiviet, D. J. *et al.* Stochasticity of metabolism and growth at the single-cell level. *Nature* **514**, 376–379 (2014).

28. Nobuyoshi Ishii, Kenji Nakahigashi, Tomoya Baba, M. R. *et al.* Multiple High-Throughput Analyses Monitor the Response of E. coli to Perturbations. *Science (80-.).* **316**, 593–598 (2007).

29. Kaczanowska, M. & Rydén-Aulin, M. Ribosome Biogenesis and the Translation Process in

Escherichia coli. *Microbiol. Mol. Biol. Rev.* **71**, 477–494 (2007).

30. Zhurinsky, J. *et al.* A coordinated global control over cellular transcription. *Curr. Biol.* **20**, 2010–2015 (2010).

31. Huls, P. G., Vischer, N. O. E. & Woldringh, C. L. Different amounts of DNA in newborn cells of Escherichia coli preclude a role for the chromosome in size control according to the 'Adder' model. *Front. Microbiol.* **9**, 1–11 (2018).

32. Hayflick, L. Biological aging is no longer an unsolved problem. *Ann. N. Y. Acad. Sci.* **1100**, 1–13 (2007).

33. Florea, M. Aging and immortality in unicellular species. *Mech. Ageing Dev.* **167**, 5–15 (2017).

34. He, S. & Sharpless, N. E. Senescence in Health and Disease. *Cell* **169**, 1000–1011 (2017).

35. Petralia, R. S., Mattson, M. P. & Yao, P. J. Aging and longevity in the simplest animals and the quest for immortality. *Ageing Res. Rev.* **16**, 66–82 (2014).

36. Bitto, A., Wang, A. M., Bennett, C. F. & Kaeberlein, M. Biochemical genetic pathways that modulate aging in multiple species. *Cold Spring Harb. Perspect. Med.* **5**, 1–26 (2015).

37. Longo, V. D., Shadel, G. S., Kaeberlein, M. & Kennedy, B. Replicative and chronological aging in saccharomyces cerevisiae. *Cell Metab.* **16**, 18–31 (2012).

38. Kaeberlein, M., Burtner, C. R. & Kennedy, B. K. Recent developments in yeast aging. *PLoS Genet.* **3**, 655–660 (2007).

39. Stewart, E. J., Madden, R., Paul, G. & Taddei, F. Aging and death in an organism that reproduces by morphologically symmetric division. *PLoS Biol.* **3**, 0295–0300 (2005).

40. Wang, P. *et al.* Robust growth of Escherichia coli. *Curr. Biol.* **20**, 1099–103 (2010).

41. Vedel, S., Nunns, H., Košmrlj, A., Semsey, S. & Trusina, A. Asymmetric Damage Segregation Constitutes an Emergent Population-Level Stress Response. *Cell Syst.* **3**, 187–198 (2016).

42. Fredriksson, Å. & Nyström, T. Conditional and replicative senescence in Escherichia coli. *Curr. Opin. Microbiol.* **9**, 612–618 (2006).

43. Coelho, M. *et al.* Fission yeast does not age under favorable conditions, but does so after stress. *Curr. Biol.* **23**, 1844–1852 (2013).

44. Kaplan, E.L; Meier, P. Nonparametric Estimation from Incomplete Observations. *J. Am. Stat. Assoc.* **53**, 457–481 (2016).

45. Symons, M. J. & Moore, D. T. Hazard rate ratio and prospective epidemiological studies. *J. Clin. Epidemiol.* **55**, 893–899 (2002).

46. Sas, A. A., Snieder, H. & Korf, J. Gompertz' survivorship law as an intrinsic principle of aging. *Med. Hypotheses* **78**, 659–663 (2012).

47. Gompertz, B. On the Nature of the Function Expressive of the Law of Human Mortality, and on a New Mode of Determining the Value of Life Contingencies. *Philos. Trans. R. Soc. London* **115**, 513–583 (1825).

48. Yang, J. S. *et al.* OASIS: Online application for the survival analysis of lifespan assays performed in aging research. *PLoS One* **6**, (2011).

49. Wang, P. *et al.* Robust growth of escherichia coli. *Curr. Biol.* **20**, 1099–1103 (2010).

50. Rojas, E., Theriot, J. A. & Huang, K. C. Response of Escherichia coli growth rate to osmotic shock. *Proc. Natl. Acad. Sci. U. S. A.* **111**, 7807–7812 (2014).

51. Boehm, A. *et al.* Genetic Manipulation of Glycogen Allocation Affects Replicative Lifespan in E. coli. *PLoS Genet.* **12**, 1–17 (2016).

52. Lele, U. N., Baig, U. I. & Watve, M. G. Phenotypic plasticity and effects of selection on cell division symmetry in escherichia coli. *PLoS One* **6**, (2011).

53. Hottes, A. K. *et al.* Bacterial Adaptation through Loss of Function. *PLoS Genet.* **9**, (2013).

54. Takeuchi, R. *et al.* Colony-live - A high-throughput method for measuring microbial colony growth kinetics - Reveals diverse growth effects of gene knockouts in Escherichia coli. *BMC Microbiol.* **14**, 1–11 (2014).

55. Baba, T. *et al.* Construction of Escherichia coli K-12 in-frame, single-gene knockout mutants: The Keio collection. *Mol. Syst. Biol.* **2**, (2006).

56. Lobritz, M. A. *et al.* Antibiotic efficacy is linked to bacterial cellular respiration. *Proc. Natl. Acad. Sci. U. S. A.* **112**, 8173–8180 (2015).

57. Coates, A., Hu, Y., Bax, R. & Page, C. The future challenges facing the development of new antimicrobial drugs. *Nat. Rev. Drug Discov.* **1**, 895–910 (2002).

58. Cardona, F., Andrés-Lacueva, C., Tulipani, S., Tinahones, F. J. & Queipo-Ortuño, M. I. Benefits of polyphenols on gut microbiota and implications in human health. *J. Nutr. Biochem.* **24**, 1415–1422 (2013).

59. Walsh, C. T. *et al.* Introduction : Antibiotic Resistance. **105**, (2005).

60. Whitesides, G. M. The origins and the future of microfluidics. *Nature* **442**, 368–373 (2006).

61. Toepke, M. W. & Beebe, D. J. PDMS absorption of small molecules and consequences in microfluidic applications. *Lab Chip* **6**, 1484–1486 (2006).

62. Chen, Q., Li, G., Nie, Y., Yao, S. & Zhao, J. Investigation and improvement of reversible microfluidic devices based on glass-PDMS-glass sandwich configuration. *Microfluid.*

Nanofluidics **16**, 83–90 (2014).

63. Iyer-biswas, S. *et al.* Scaling laws governing stochastic growth and division of single bacterial cells. (2014) doi:10.1073/pnas.1403232111.

64. Susman, L. *et al.* Individuality and slow dynamics in bacterial growth homeostasis. *Proc. Natl. Acad. Sci. U. S. A.* **115**, E5679–E5687 (2018).

65. Yang, D., Jennings, A. D., Borrego, E., Retterer, S. T. & Männik, J. Analysis of Factors Limiting Bacterial Growth in PDMS Mother Machine Devices. *Front. Microbiol.* **9**, 871 (2018).

66. Rohm and Haas Electronic Materials. MICROPOSIT S1800 G2 SERIES: For Microlithography Applications. 1–6 (2006).

67. MicroChem. SU-8 2000.5-2015 Permanent Epoxy Negative Photoresist PROCESSING GUIDELINES. *SU-8 2000 Datasheet* 1–5 (2015).

68. Kohram, M., Vashistha, H., Leibler, S., Xue, B. & Salman, H. Bacterial Growth Control Mechanisms Inferred from Multivariate Statistical Analysis of Single-Cell Measurements. *Curr. Biol.* 1–10 (2020) doi:https://doi.org/10.1016/j.cub.2020.11.063.

69. Ferrezuelo, F. *et al.* The critical size is set at a single-cell level by growth rate to attain homeostasis and adaptation. *Nat. Commun.* **3**, 1012 (2012).

70. Tzur, A., Kafri, R., LeBleu, V. S., Lahav, G. & Kirschner, M. W. Cell growth and size homeostasis in proliferating animal cells. *Science* **325**, 167–71 (2009).

71. Bonny, M., Fischer-Friedrich, E., Loose, M., Schwille, P. & Kruse, K. Membrane Binding of MinE Allows for a Comprehensive Description of Min-Protein Pattern Formation. *PLoS Comput. Biol.* **9**, e1003347 (2013).

72. Banerjee, S. *et al.* Biphasic growth dynamics control cell division in Caulobacter crescentus. *Nat. Microbiol.* **2**, 17116 (2017).

73. Osella, M., Nugent, E. & Cosentino Lagomarsino, M. Concerted control of Escherichia coli cell division. *Proc. Natl. Acad. Sci. U. S. A.* **111**, 3431–5 (2014).

74. Männik, J. *et al.* Robustness and accuracy of cell division in Escherichia coli in diverse cell shapes. *Proc. Natl. Acad. Sci. U. S. A.* **109**, 6957–62 (2012).

75. Grilli, J., Osella, M., Kennard, A. S. & Lagomarsino, M. C. Relevant parameters in models of cell division control. *Phys. Rev. E* **95**, (2017).

76. Kennard, A. S. *et al.* Individuality and universality in the growth-division laws of single E. Coli cells. *Phys. Rev. E - Stat. Nonlinear, Soft Matter Phys.* **93**, 012408 (2016).

77. Brenner, N. *et al.* Single-cell protein dynamics reproduce universal fluctuations in cell

populations. *Eur. Phys. J. E. Soft Matter* **38**, 102 (2015).

78. Lutz, R. & Bujard, H. Independent and tight regulation of transcriptional units in Escherichia coli via the LacR/O, the TetR/O and AraC/I1-I2 regulatory elements. *Nucleic Acids Res.* **25**, 1203–10 (1997).

79. Klein, J. *et al.* TLM-Tracker: software for cell segmentation, tracking and lineage analysis in time-lapse microscopy movies. *Bioinformatics* **28**, 2276–2277 (2012).

80. Freedman, D. *Statistical Models: Theory and Practice*. (cambridge university press, 2009).

81. Hidalgo, B. & Goodman, M. Multivariate or multivariable regression? *Am. J. Public Health* **103**, 39–40 (2013).

82. Willis, L. *et al.* Cell size and growth regulation in the Arabidopsis thaliana apical stem cell niche. *Proc. Natl. Acad. Sci. U. S. A.* **113**, E8238–E8246 (2016).

83. Tanouchi, Y. *et al.* A noisy linear map underlies oscillations in cell size and gene expression in bacteria. *Nature* **523**, 357–360 (2015).

84. Sandler, O. *et al.* Lineage correlations of single cell division time as a probe of cell-cycle dynamics. *Nature* **519**, 468–471 (2015).

85. Elowitz, M. B., Levine, A. J., Siggia, E. D. & Swain, P. S. Stochastic gene expression in a single cell. *Science (80-.).* **297**, 1183–1186 (2002).

86. Raser, J. M. & O'Shea, E. K. Control of stochasticity in eukaryotic gene expression. *Science* **304**, 1811–4 (2004).

87. Ozbudak, E. M., Thattai, M., Kurtser, I., Grossman, A. D. & van Oudenaarden, A. Regulation of noise in the expression of a single gene. *Nat. Genet.* **31**, 69–73 (2002).

88. You, C. *et al.* Coordination of bacterial proteome with metabolism by cyclic AMP signalling. *Nature* **500**, 301–306 (2013).

89. Brenner, N. *et al.* Universal protein distributions in a model of cell growth and division. *Phys. Rev. E - Stat. Nonlinear, Soft Matter Phys.* **92**, 1–6 (2015).

90. Schmoller, K. M., Turner, J. J., Kõivomägi, M. & Skotheim, J. M. Dilution of the cell cycle inhibitor Whi5 controls budding-yeast cell size. *Nature* **526**, 268–272 (2015).

91. Iyer-Biswas, S., Crooks, G. E., Scherer, N. F. & Dinner, A. R. Universality in Stochastic Exponential Growth. *Phys. Rev. Lett.* **113**, 028101 (2014).

92. Di Talia, S., Skotheim, J. M., Bean, J. M., Siggia, E. D. & Cross, F. R. The effects of molecular noise and size control on variability in the budding yeast cell cycle. *Nature* **448**, 947–51 (2007).

93. Soifer, I., Robert, L. & Amir, A. Single-cell analysis of growth in budding yeast and bacteria

reveals a common size regulation strategy. *Curr. Biol.* **26**, 356–361 (2016).

94. Campos, M. *et al.* A constant size extension drives bacterial cell size homeostasis. *Cell* **159**, 1433–46 (2014).

95. Klumpp, S. & Hwa, T. Growth-rate-dependent partitioning of RNA polymerases in bacteria. *Proc. Natl. Acad. Sci. U. S. A.* **105**, 20245–20250 (2008).

96. Travers, A. & Muskhelishvili, G. DNA supercoiling - A global transcriptional regulator for enterobacterial growth? *Nat. Rev. Microbiol.* **3**, 157–169 (2005).

97. Rasmussen, A. A. *et al.* Regulation of ompA mRNA stability: The role of a small regulatory RNA in growth phase-dependent control. *Mol. Microbiol.* **58**, 1421–1429 (2005).

98. Kubista, M., Akerman, B. & Norden, B. Characterization of Interaction between DNA and 4',6-Diamidino-2-phenylindole by Optical Spectroscopy. *Biochemistry* **26**, 4545–4553 (1987).

99. Chao, Y. & Zhang, T. Optimization of fixation methods for observation of bacterial cell morphology and surface ultrastructures by atomic force microscopy. *Appl. Microbiol. Biotechnol.* **92**, 381–392 (2011).

100. Vashistha, H., Kohram, M. & Salman, H. Non-genetic inheritance restraint of cell-to-cell variation. *arXiv* (2020).

101. Clark, D. J. Regulation of deoxyribonucleic acid replication and cell division in Escherichia coli B-r. *J. Bacteriol.* **96**, 1214–1224 (1968).

102. Eberhardt, C., Kuerschner, L. & Weiss, D. S. Probing the catalytic activity of a cell division-specific transpeptidase in vivo with β-lactams. *J. Bacteriol.* **185**, 3726–3734 (2003).

103. Boyle, J. V., Cook, T. M. & Goss, W. A. Mechanism of action of nalidixic acid on Escherichia coli. Vi. Cell-free studies. *J. Bacteriol.* **97**, 230–236 (1969).

104. Nadkarni, S. R., Reddy, G. C. S., Patel, M. V, Reden, J. & Ganguli, B. N. IN VITRO SYNERGY OF ACTITHIAZIC ACID, AN ANTIMETABOLITE OF BIOTIN, COMBINED WITH CEPHALEXIN. *J. Antibiot. (Tokyo).* **XXXVI**, 1567–1568 (1983).

105. D'Ari, R. & Nanninga, N. Cell shape and division in escherichia coli. *Res. Microbiol.* **141**, 9–16 (1990).

106. Strogatz, S. H. *Nonlinear Dynamics and Chaos : With Applications to Physics, Biology, Chemistry, and Engineering*.

107. Amir, A., Babaeipour, F., McIntosh, D. B., Nelson, D. R. & Jun, S. Bending forces plastically deform growing bacterial cell walls. *Proc. Natl. Acad. Sci. U. S. A.* **111**, 5778–5783 (2014).

108. Berg, H. C. Motile behavior of bacteria. *Phys. Today* **53**, 24–29 (2000).

109. Smith, A. *et al.* The culture environment influences both gene regulation and phenotypic heterogeneity in Escherichia coli. *Front. Microbiol.* **9**, 1–13 (2018).

110. Orth, D. S., Lutes, C. M. & Smith, D. K. Effect of culture conditions and method of inoculum preparation on the kinetics of bacterial death during preservative efficacy testing. *J. Soc. Cosmet. Chem.* **40**, 193–204 (1989).

111. Sourjik, V. & Wingreen, N. S. Responding to chemical gradients: Bacterial chemotaxis. *Curr. Opin. Cell Biol.* **24**, 262–268 (2012).

112. Demir, M. & Salman, H. Bacterial thermotaxis by speed modulation. *Biophys. J.* **103**, 1683–1690 (2012).

113. Sprenger, W. W., Hoff, W. D., Armitage, J. P. & Hellingwerf, K. J. The eubacterium Ectothiorhodospira halophila is negatively phototactic, with a wavelength dependence that fits the absorption spectrum of the photoactive yellow protein. *J. Bacteriol.* **175**, 3096–3104 (1993).

114. Robinson, K. R. The responses of cells to electrical fields: A review. *J. Cell Biol.* **101**, 2023–2027 (1985).

115. Şimşek, E. & Kim, M. The emergence of metabolic heterogeneity and diverse growth responses in isogenic bacterial cells. *ISME J.* **12**, 1199–1209 (2018).

116. Hadizadeh Yazdi, N., Guet, C. C., Johnson, R. C. & Marko, J. F. Variation of the folding and dynamics of the Escherichia coli chromosome with growth conditions. *Mol. Microbiol.* **86**, 1318–1333 (2012).

117. Lobritz, M. A. *et al.* Antibiotic efficacy is linked to bacterial cellular respiration. *Proc. Natl. Acad. Sci. U. S. A.* **112**, 8173–8180 (2015).

118. Colville, K., Tompkins, N., Rutenberg, A. D. & Jericho, M. H. Effects of poly(L-lysine) substrates on attached escherichia coli bacteria. *Langmuir* **26**, 2639–2644 (2010).

119. Klein, J. *et al.* TLM-tracker: Software for cell segmentation, tracking and lineage analysis in time-lapse microscopy movies. *Bioinformatics* **28**, 2276–2277 (2012).

120. Beg, Q. K. *et al.* Intracellular crowding defines the mode and sequence of substrate uptake by Escherichia coli and constrains its metabolic activity. *Proc. Natl. Acad. Sci. U. S. A.* **104**, 12663–12668 (2007).

121. Zhou, Y. *et al.* Carbon catabolite repression correlates with the maintenance of near invariant molecular crowding in proliferating E. coli cells. *BMC Syst. Biol.* **7**, (2013).

122. Kotte, O., Volkmer, B., Radzikowski, J. L. & Heinemann, M. Phenotypic bistability in Escherichia coli 's central carbon metabolism . *Mol. Syst. Biol.* **10**, 736 (2014).

123. Raj, A. & van Oudenaarden, A. Nature, Nurture, or Chance: Stochastic Gene Expression and Its Consequences. *Cell* **135**, 216–226 (2008).

124. Balaban, N. Q., Merrin, J., Chait, R., Kowalik, L. & Leibler, S. Bacterial persistence as a phenotypic switch. *Science (80-.).* **305**, 1622–1625 (2004).

125. Yamaichi, Y., Fogel, M. A. & Waldor, M. K. par genes and the pathology of chromosome loss in Vibrio cholerae. *Proc. Natl. Acad. Sci. U. S. A.* **104**, 630–635 (2007).

126. Yamamoto, N. *et al.* Update on the Keio collection of Escherichia coli single-gene deletion mutants. *Mol. Syst. Biol.* **5**, (2009).

127. Cooper, S. Leucine uptake and protein synthesis are exponential during the division cycle of Escherichia coli B/r. *J. Bacteriol.* **170**, 436–438 (1988).

128. Lutz, R. & Bujard, H. Independent and tight regulation of transcriptional units in escherichia coli via the LacR/O, the TetR/O and AraC/I1-I2 regulatory elements. *Nucleic Acids Res.* **25**, 1203–1210 (1997).

129. Skoko, D. *et al.* Mechanism of Chromosome Compaction and Looping by the Escherichia coli Nucleoid Protein Fis. *J. Mol. Biol.* **364**, 777–798 (2006).

130. Sauls, J. T., Li, D. & Jun, S. Adder and a coarse-grained approach to cell size homeostasis in bacteria John. *Curr Opin Cell Biol.* **38**, 38–44 (2017).

131. Hashimoto, M. *et al.* Noise-driven growth rate gain in clonal cellular populations. *Proc. Natl. Acad. Sci.* **113**, 3251–3256 (2016).

132. Brunk, U. T. & Terman, A. The mitochondrial-lysosomal axis theory of aging: Accumulation of damaged mitochondria as a result of imperfect autophagocytosis. *Eur. J. Biochem.* **269**, 1996–2002 (2002).

133. Burhans, W. C. & Weinberger, M. DNA replication stress, genome instability and aging. *Nucleic Acids Res.* **35**, 7545–7556 (2007).

134. Lin, J., Epel, E. & Blackburn, E. Telomeres and lifestyle factors: Roles in cellular aging. *Mutat. Res. - Fundam. Mol. Mech. Mutagen.* **730**, 85–89 (2012).

135. Harris, C. R. *et al.* Array programming with NumPy. *Nature* **585**, 357–362 (2020).

136. Inoue, T. *et al.* Genome-wide screening of genes required for swarming motility in Escherichia coli K-12. *J. Bacteriol.* **189**, 950–957 (2007).

137. Keseler, I. M. *et al.* The EcoCyc database: Reflecting new knowledge about Escherichia coli K-12. *Nucleic Acids Res.* **45**, D543–D550 (2017).

138. P.D. Karp, D. Weaver, S. Paley, C. Fulcher, A. Kubo, A. Kothari, M. Krummenacker, P. Subhraveti, D. Weerasinghe, S.Gama-Castro, A.M. Huerta, L., Muniz-Rascado,

C.Bonavides-Martinez, V.Weiss, M.Peralta-Gil, A. Santos-Zavaleta, I. Schroder, A. Mackie, R. Gu, and I. P. EcoCyc : Escherichia coli K-12 substr . MG1655 Cellular Overview - EcoCyc database. *EcoSal Plus* (2014) doi:10.1128/ecosalplus.ESP-0009-2013.